AB

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/90, 15/61

A1

(11) International Publication Number: WO 94/09124

(43) International Publication Date: 28 April 1994 (28.04.94)

(21) International Application Number: PCT/NO93/00151

(22) International Filing Date: 8 October 1993 (08.10.93)

(30) Priority data:
9221163.0 8 October 1992 (08.10.92) GB

(71) Applicants *(for all designated States except US)*: PRONOVA BIOPOLYMER A.S [NO/NO]; Tomtegt. 36, N-3000 Drammen (NO). NOBIPOL [NO/NO]; Sem Sælandsvn. 6/8, N-7034 NTH Trondheim (NO).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : ERTESVÅG, Helga [NO/NO]; Blåklokkeveien 29D, N-7035 Trondheim (NO). VALLA, Svein [NO/NO]; Naustanveien 2, N-7560 Vikhamar (NO). SKJÅK-BRÆK, Gudmund [NO/NO]; Nedre Bergsvingen 6, N-7016 Trondheim (NO). LARSEN, Bjørn [NO/NO]; Egganveien 16, N-7081 Trondheim (NO).

(74) Agent: SANDBU, Dahl, Elisabeth; Pronova a.s, Strandveien 5, N-1324 Lysaker (NO).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DNA COMPOUNDS COMPRISING SEQUENCES ENCODING MANNURONAN C-5-EPIMERASE

(57) Abstract

DNA compounds encompassing sequences coding for enzymes having mannuronan C-5-epimerase activity are disclosed and a process for the preparation of such enzymes. The genetic sequences and enzymes prepared may be used in the production of alginates having a definite G/M ratio and block structure. Alginates having a definite G/M ratio may also be produced by selective inactivation of the genetic sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic			SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovak Republic
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	UA	Ukraine
DE	Germany	MG	Madagascar	US	United States of America
DK	Denmark	ML	Mali	UZ	Uzbekistan
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

DNA COMPOUNDS COMPRISING SEQUENCES ENCODING MANNURONAN C-5-EPIMERASE

The present invention concerns DNA compounds encompassing sequences coding for enzymes having mannuronan C-5-epimerase activity, a process for the preparation of such enzymes, the use of said genetic sequences in production of alginates having a definite G/M ratio and block structure, and the production of alginates having a definite G/M ratio by inactivating said genetic sequences.

Throughout this application, reference is made to publications from the scientific and patent literature. Publications so cited are hereby incorporated in their entirety by reference.

In this application the term gene is used to indicate a genetic sequence which encodes a protein, independent of whether the protein encoded by this genetic sequence is expressed or not in the natural host organism under those conditions.

Alginates are a family of polysaccharides, which are synthesized in brown algae as well as in bacteria, such as Azotobacter vinelandii and Azotobacter chroococcum. Alginates are also synthesized by some strains of Pseudomonas sp.

Chemically, alginates are unbranched binary copolymers of 1-4 linked ß-D-mannuronic acid, termed M, and its C-5 epimer α-L-guluronic acid, termed G.

Alginates derived from seaweeds and Azotobacter are generally true block copolymers where the monomers are arranged in homopolymeric stretches of M, termed M blocks, and homopolymeric stretches of G, termed G blocks, interspaced with regions containing both monomers, normally termed alternating blocks or MG blocks. The composition and sequential structure of alginates vary widely depending on

the source. Alginates produced by Pseudomonas, however, do not have any G blocks.

Several functional properties such as the capacity to form gels and the binding of water depend on the M/G ratio and on the length of the various blocks. A relatively high content of G blocks, for instance gives good gelling properties, due to ionic cross linking of chains which takes place when Ca^{2+}-ions are added to an alginate solution. The composition and block structure also influence on the immunological properties of alginates. [Otterlei et al, J.of Immunotherapy 10, 286-291, (1991)] have shown that alginates with a high content of mannuronic acid blocks are very potent nontoxic immunostimulants.

At present industrial production of alginates rely exclusively on algal sources. The range in composition is however limited as the highest content of guluronic acid to be found is 75% and the lowest 25%. Furthermore there are no suitable sources for alginate with a G content in the range of 42-54%. In the field of biotechnology or biomedicine, polymers with extreme compositions, such as a high G for immobilization of cells, [Martinsen A., Skjåk-Bræk G. and Smidsrød O., Biotechnol. Bioeng. 33, 79-86, (1989)] and a high M (90-100%) as immunostimulants [Otterlei et al, J.of Immunotherapy 10, 286-291, (1991)] are of major interest.

The key enzyme responsible for generation of the G blocks is called mannuronan C-5-epimerase. It was previously thought that only one enzyme having a certain amino acid sequence would exhibit this activity. It has now surprisingly been found that there exist at least five genes encoding enzymes having this activity. Some of these enzymes differ in molecular weight and amino acid sequence. The genes were found adjacent to each other in the bacterium Azotobacter vinelandii. It has also been found

that the amino acid sequence of the enzyme affects the activity of the enzyme, not only in terms of potency but also in the type of alginate formed, for example, altering the content of guluronic acid and the single/block G content of the alginate.

In [Larsen, B. and Haug, A., Carbohydr. Res. 17, (1971), 287-296 and 297-308] the isolation of mannuronan C-5-epimerase from liquid cultures of Azotobacter vinelandii is reported. In the following, this epimerase will be termed mannuronan C-5-epimerase (2), and the DNA sequence encoding for it will correspondingly be denominated E2.

In [Skjåk-Bræk, G and Larsen, B, Carbohydr. Res. 103:133-136, (1982)], the purification of mannuronan C-5-epimerase (2) by affinity chromatography on alginate sepharose is disclosed. In a separate paper [Skjåk-Bræk, G and Larsen, B., Carbohydrate Research, 139, (1985) 273-283] the characterization of this enzyme is disclosed. Further, the activity of the enzyme is described as an ability to epimerize both bacterial and algal alginate having a wide range in monomer composition and sequence of units.

From PCT/WO 86/03781 and Japanese Patent Application J63233797 it is known to produce alginic acid and/or alginate having a high content of guluronic acid by action of the enzyme (E2) on an alginic acid or alginate, whereby the G content increases.

In [Chitnis, C.E. and Ohman, D.E., J. Bacteriol., 172, p2894-2900, (1990)] the gene sequences involved in the introduction of guluronic acid into exopolysaccharides from Pseudomonas aeruginosa have been reported. However, the nature of the enzyme responsible for this process has not been identified. Since this genus of bacteria is unable to produce alginate containing G blocks [Skjåk-Bræk, G.,

Larsen, B. and Grasdalen, H. Carbohydr. Res. 54 (1986) 169-174] it is believed that the epimerization system in alginate producing Pseudomonas is fundamentally different from the epimerase in brown algae and in Azotobacter vinelandii. It seems likely that the Pseudomonas enzyme is a monomer epimerase acting at the sugar nucleotide level, and as such is unable to introduce G-blocks into already polymerized alginates.

Production of mannuronan C-5-epimerase from Azotobacter vinelandii culture is difficult due to a very low yield. It is also a major obstacle that the enzyme is secreted together with copious amounts of highly viscous alginate which hampers the purification of the enzyme. Although alginates are secreted by some bacteria, an industrial production based on these microorganisms has not been successful. The main reasons are due to the difficulties in controlling the composition and molecular size of the exopolysaccharides. The content of guluronic acid blocks in the alginate from Azotobacter vinelandii tends to be too low for making a polymer with good gelling properties.

Alginates with a high M content having immunogenic properties as reported above, are produced by Pseudomonas aeruginosa, but this organism is unattractive from a production point of view, as it is unstable in the production of the polymer. Further, the organism is known to be a secondary pathogen in patients suffering from cystic fibrosis.

Thus, in order to produce medical grade alginates with a defined monomer composition and sequential structure there is a need for improved methods for controlling the biosynthesis of alginate, through controlling the key enzyme, the mannuronan C-5-epimerase.

The present invention is directed to cloned DNA fragments

encoding mannuronan C-5-epimerase. The invention is encompassed by vectors which contain DNA fragments encoding mannuronan C-5-epimerase linked to DNA elements which direct the expression of mannuronan C-5-epimerase from the cloned DNA encoding the protein. The invention also provides for microorganisms which express the mannuronan C-5-epimerase protein from the cloned DNA as a source of the purified protein and also as a source of alginates of altered composition. Strains in which the expression level of the mannuronan C-5-epimerase gene is altered or in which one, several or all of the mannuronan C-5-epimerase genes have been inactivated are also within the scope of the present invention. The invention further encompasses methods for producing alginates either very efficiently, or having altered composition, or both, by culturing microorganisms having altered levels of expression of a mannuronan C-5-epimerase gene.

The invention further features selection of epimerase to achieve a desired level of guluronic acid, and alter the single/block G characteristics of the enzyme. In a further embodiment, the invention features the production of synthetic proteins and DNA encoding such proteins which have mannuronan C-5-epimerase activity.

Brief Description of Figures

Figure 1 shows the amino acid sequence of the N-terminal end of the 122 kd protein, and the nucleotide sequence of the corresponding oligonucleotide. The DNA probe was synthesized as a mixture (in equal ratios) of the 64 possible combinations that could be deduced from the first seven amino acids in the sequence of the 122 kd protein. N indicates that all four bases were used at this position.

Figure 2 is a restriction endonuclase map of the combined inserts in plasmids pHE14, pHE16, pBD1, pHE18 and pML1.

The numbers at the bottom line indicate the molecular sizes in bp. The arrow indicates the localization and orientation of the sequence homologous to the synthetic oligonucleotide used for screening the library. The five genes (open reading frames) found by sequencing are marked by boxes and denoted E4, E1, E2, E3 and E5. E1 corresponds to Epimerase I.

Figure 3 shows Mannuronan C-5-epimerase (1) activity of a portion of the E1 encoded protein as a function of cell growth. *: OD_{600} of cell culture. o: Epimerase activity given as dpm/ml of cell culture. The strain used in this experiment was DH5α(pHE5), and the extracts were incubated with the substrate for 23 hours.

Figure 4 shows the kinetics of ^{3}H release. The enzyme activity was assayed by using an extract prepared from IPTG-induced JM105 cells containing pHE5 (see legend to Table 3).

Figure 5 shows the homologies between and within the different genes. Boxes with the same letter are homologous to each other. Gaps are introduced to optimize the alignment. E1-E4 are defined as appears from Figures 2 and 6.

Figure 6 shows the nucleotide sequences and corresponding amino acid sequences for E4, E1, E2 and part of E3.

Figure 7 shows the alignment of the DNA sequences of the A blocks from E4, E1, E2 and E3.

Figure 8 shows the alignment of the deduced amino acid sequences of the A blocks from E4, E1, E2 and E3.

Figure 9 shows the alignment of the DNA sequences of the R blocks from E4, E1, E2 and E3.

Figure 10 shows the alignment of the deduced amino acid sequences from the R blocks of E4, E1, E2 and E3.

Figure 11 shows ^{1}H-NMR spectra of alginate epimerized by extracts from A: DH5α(pHE8) (truncated epimerase 1); B: JM109(pBD9); C: no extract. The peak to the left gives the signal from G-1; the peak in the centre gives the combined signal from GM-5 and M-1 and the peak to the right gives the signal from GG-5.

Figure 12 shows the nucleotide sequence and corresponding amino acid sequence of E2.

Now according to the present invention genetic sequences have been found which encode enzymes having mannuronan C-5-epimerase activity, and thus the first aspect of the invention is pure isolated DNA comprising nucleotide sequences encoding mannuronan C-5-epimerase activity.

The sequence of amino acids proximal to the amino terminus of purified mannuronan C-5-epimerase protein was determined [G. Skjåk Bræk et al., Carbohydr. Res. 103:133-136 (1982)]. This data was used to derive a sequence for an oligonucleotide probe which was used to screen a gene library of Azotobacter vinelandii DNA. One result of this screening experiment was the surprising discovery of a second gene and thereafter three further genes including at least one genetic block A were found. Thus, altogether there appear to be at least five different genes encoding proteins having mannuronan C-5-epimerase activity. Accordingly, it is a second object of the present invention to provide for alternative DNA sequences encoding mannuronan C-5-epimerase.

Three different blocks of genetic sequences, designated A, R and S, are found in the genes. These genetic blocks are most commonly found in combinations wherein the A appears

one time or two times, the R block appears from 0 to at least 5 times and the S block appears from 0-1 time.

There is a high degree of consensus in the nucleotide sequences of each block for the different genes (1-5). Accordingly, it is a third object of the present invention to provide for DNA sequences encoding mannuronan C-5-epimerase and comprising the DNA blocks A and/or S and/or R, wherein A may appear more than once and R if present may appear singly or in repeats of up to at least 5 or 6 times.

The consecutive order of the three blocks if all three blocks are present, is preferably A, R and S. However, it has been shown that a reversed consecutive order, wherein for instance R appears before A also gives a gene encoding a mannuronan C-5-epimerase. Thus, the invention further encompasses genetic sequences having any order and any number of the blocks A, R and S.

Another aspect of the present invention concerns the use of said genetic sequences for the preparation of the mannuronan C-5-epimerase in recombinant host cells. It is especially preferred to insert the gene into hosts such as bacteria, for instance Escherichia coli or Bacillus subtilis or in yeast. The cloning and expression of the genetic sequence as described above in E. coli is described in the Examples.

The present invention also encompasses recombinant expression plasmids that can be used to produce the mannuronan C-5-epimerase proteins in a host microorganism. Such expression plasmids are made by inserting a DNA fragment encoding mannuronan C-5-epimerase into a vector which contains appropriate expression elements, such as (but not limited to) a promoter, ribosome binding site, translational initiation site and transcription terminator. The expression plasmids can be adapted for transformation

into many different commonly used host organisms in which it might be desired to produce the mannuronan C-5-epimerase.

The techniques for insertion of foreign genes into commonly employed hosts are known in the art, as described for instance in [METHODS IN ENZYMOLOGY, Vol. 185, Gene Expression Technology, Ed. D.V.Goeddel, Academic Press, Inc. (1990)]. Further by choice of a broad host range vector and a suitable promoter as known in the art, and described for instance in [J.L.Ramos et al, FEBS Letters, Vol. 226, 2, 241-246] it will be possible to insert and express the mannuronan C-5-epimerase genetic sequences in many different hosts.

This will make possible the production of large quantities of one or all of the pure enzymes having this activity, while avoiding the problems of separating the enzymes from the alginate.

By inserting a high copy-number vector comprising the genetic sequences encoding the epimerase into a natural alginate producing bacterium such as Azotobacter vinelandii an enhanced production of the enzymes would be possible.

Over expression of the epimerases in a natural host could also be achieved by using a promoter which drives high-level expression of the enzymes. By blocking other genetic sequences coding for the alginate production, the production of pure enzymes may be achieved.

Yet another aspect of the invention is the selective inactivation of the mannuronan C-5-epimerase genes in the natural host organism so as to provide for bacterial production of alginates having a low content of G blocks or even a pure poly-M alginate. This is accomplished by inserting nucleotides into one, several or all of the

mannuronan C-5-epimerase genes in the natural host organism Azotobacter. It is especially preferred to insert a DNA fragment encoding a selectable marker gene, preferably a gene conferring antibiotic resistance. Insertion of a selectable marker allows selection of those bacteria in which the insertion has been successfully accomplished. By choosing different selectable markers, for example providing resistance to different antibiotics, it is possible to select recombinants that have incorporated inserted sequences into some or all of the mannuronan C-5-epimerase genes. Thus, selective production of bacterial strains in which one of the mannuronan C-5-epimerase genes, several or all of them have been inactivated is possible.

A second method of inactivating all of the epimerase genes is to transform a cell of the natural host strain, Azotobacter with a vector which expresses an antisense RNA which specifically binds to mRNA transcribed from these genes. Use of promoters of varying strength to drive expression of the antisense RNA in the creation of the vectors used to transform the cells allows production of strains having varying ratios of G-blocks to M-blocks in the alginate produced. Use of inducible promoters to drive expression of the antisense RNA allows the creation of strains which can produce alginates of variable composition, depending on culture conditions. Clearly, if the recombinant host organism is the natural host, Azotobacter, it is possible to enhance production of one epimerase gene while leaving expression of the others at their normal level, thus producing a strain which makes an alginate having an altered ratio of G blocks to M blocks. A strain which makes alginate having 0-25% M blocks is preferred.

Alternatively, all but one of the epimerase genes can be inactivated, as described above, and the expression of the remaining epimerase gene can be controlled by a regulated

promoter. A strain carrying such a complement of epimerase genes would thus produce alginates having a high content of G-blocks, especially from 75-98%. Another means for making a strain for producing alginates having a high G-block content is to inactivate all but one of the mannuronan C-5-epimerase genes by insertion and control the remaining gene by antisense RNA, using an inducible promoter to regulate transcription of the antisense RNA gene. A still further means for making a strain for producing alginates having a high G-block content is by inactivating all naturally occurring genes and introducing a regulated gene through a vector.

Thus the present invention also includes a process for the construction of a recombinant host cell capable of expressing mannuronan C-5-epimerase activity by transforming said host cell with a recombinant DNA expression vector that comprises: (a) a promoter and translational activating sequence that function in said host cell; and (b) a DNA sequence encoding mannuronan C-5-epimerase comprising at least a DNA block A and/or a DNA block S and/or a DNA block R, positioned for expression from said promoter and translational activity sequence.

Also the present invention encompasses a process for the bacterial production of pure poly-M alginate or tailored alginates having a lower G block content, preferably in the range from 0-25%, by blocking the DNA sequences encoding the enzymes in a natural host by insertion of a foreign genetic sequence into one, several or all genetic sequences encoding mannuronan C-5-epimerase.

Other methods for achieving the same end will be known for persons skilled in the art and are hereby included into the scope of the present invention.

A further aspect of the invention are the novel enzymes

having mannuronan C-5-epimerase activity. The amino acid sequences and their degree of homology will appear from Figures 6-11.

Also as known by a person skilled in the art, variations in the nucleotide sequence which nevertheless encode proteins having the same activity as the wild-type mannuronan C-5-epimerase are encompassed within this invention.

Variations within the amino acid sequence may also encompass deletions, substitutions and additions which do not substantially change the biological activity.

Also, it is possible to make a synthetic DNA sequence encoding a mannuronan C-5-epimerase by techniques well known in the art. See for instance, [Itakura et al., Science 198:1056 (1977)] and [Crea et al. (Proc. Natl. Acad. Sci. USA 75:5765 (1978)] and also U.S. Patents 4,800,159 and 4,683,202 and also published European patent application EP-A-0258017. Synthetic enzymes may be made by incorporating different combinations of the A, R and S elements, to maintain epimerase activity. The resultant alginate composition can be varied by enzyme selection.

Materials and General Methods

Bacterial strains, plasmids, and phage. Strains, plasmids, and phages are listed in Table 1.

The bacterial strain of A. vinelandii used in these experiments, is freely available from Bjørn Larsen, Inst. of Biotechnology, Lab. for Marine Biochemistry, 7034 Trondheim - NTH, Norway or Svein Valla, Unigen, Center for Molecular Biology, University of Trondheim, 7005 Trondheim, Norway and has been deposited at the Belgian Coordinated Collections of Microorganisms (BCCM) at the Laboratorium voor Microbiologie (LMG) at Universiteit Gent (RUG), K.L.

Ledeganckstraat 35, B-9000 Gent. Other strains of A.vinelandii mentioned in Example 9 have the following ATCC numbers: ATCC 478, ATCC 12837 and ATCC 12518. Plasmids/ strains DH5α(pHE14), JM109(pHE16), JM109(pBD1), JM109(pHE18) and SURE^{TM}(pML1) have been deposited at BCCM at the Laboratorium voor Moleculaire Biologie (LMBP) (same address as LMG) and have the following numbers

Growth of bacteria and phages. A. vinelandii was grown at 30°C with shaking in a nitrogen-free medium (9.8 mM K_2HPO_4/KH_2PO_4, 0.8 mM $MgSO_47H_2O$, 3.4 mM NaCl, 0.34 mM $CaCl_2$, 8.7 μM $Na_2MoO_42H_2O$, 54 μM $FeSO_47H_2O$, 1% sorbitol, pH 7.4). E. coli was grown in LB-medium [Sambrook J, Fritsch, E.F. and Maniatis T., Molecular cloning, A laboratory manual, 2nd ed., Cold Spring Harbour Laboratory Press, New York, (1989)] with shaking at 37°C. When the cells were to be used for growth of phages the LB-medium was supplemented with 2.5 mM $CaCl_2$, 10 mM $MgCl_2$, and 0.4% maltose. Phages were plated on strain Q359 on L-agar (LB-medium supplemented with 2% agar). Phage LB-medium supplemented with either 0.8% agar (titrations and gene library amplification) or 0.8% agarose (screening of gene library and preparation of phage lysates) was used for overlaying agar.

Standard recombinant DNA technology. Restriction endonuclease digestions, removal of cohesive DNA ends by using the 3′ exonuclease activity of T4 DNA polymerase, ligations, agarose gel electrophoresis, and end-labelling with ^{32}P were performed according to standard protocols [Sambrook J, Fritsch, E.F. and Maniatis T ., Molecular cloning, A laboratory manual, 2nd ed., Cold Spring Harbour Laboratory Press, New York, (1989)]. Transformations were performed as described by [Chung, C.T., Niemela S.L. and Miller R.H., Proc. Natl. Acad. Sci USA, 86, 2172-2175, (1989)], and DNA sequencing was performed according to [Sanger F., Nicklen S., and Coulsom,. A.R., Proc. Natl.

Acad. Sci USA, 74, 4563 (1977)].

Viscosimetric measurement. The alginate used in this experiment was obtained from Ascophyllum nodosum and had an intrinsic viscosity in 0.1 M NaCl of 17.6 dl/g at 25°C. The viscosity was determined by an Ubbelhode viscosimeter.

NMR spectroscopy. The substrate used in these analyses was a low guluronic acid-containing alginate obtained from the brown algae Ascophyllum nodosum, and was prepared as described previously [Larsen, B., Proceedings of the Tenth International Seaweed Symposium, Ed: Levring, T.Gothenburg, p7-33, (1980)]. For the NMR analyses epimerase was obtained from IPTG-induced E. coli JM105 cells containing pHE5. 250 ml cell culture were harvested by centrifugation and resuspended in 20 ml of 10 mM Tris, 0.34 mM $CaCl_2$, pH 7.0. After ultrasonication, the solution was centrifuged at 31.000 x g for 1 hour. The supernatant was stored frozen at 70°C. After thawing the supernatant was filtered through a membrane with pore size 0.22 μm, and the enzyme was further purified on a Mono Q HR515 (Pharmacia) ion exchange column. The enzyme was eluted with a 0-1 M NaCl salt gradient (in the same buffer as the applied solution), and was collected in 2 ml at approximately 0.6 M NaCl. To each of two tubes was added 0.28 ml of this enzyme solution (0.9 mg/ml total protein), 1 ml alginate (7.5 mg/ml in H_2O), and 4.62 ml 2,3,6-trimethylpyridine buffer (see above). $CaCl_2$ was then added to a total reaction volume of 6 ml such that one tube contained 0.85 mM, and one contained 3.4 mM $CaCl_2$. After incubation at 30°C for 20 hours Na_2EDTA (10 mM) was added to chelate the Ca^{2+}-ions, and the solutions were then dialyzed extensively against distilled water. The dialyzed alginate solutions were freeze-dried and then dissolved in D_2O. NMR spectroscopy of these solutions were finally performed according to [Grasdalen H., Larsen B., and Smidsrød O., Carbohydr. Res., 68, 23-31 (1979)] (Table 4). Further analysis was carried out in a similar fashion for

DH5α(pHE8), JM109(pHE16) and JM109(pBD9). The results in Table 4 conclusively demonstrate that the enzymatic activity is mannuronan C-5-epimerase activity. This activity is expressed from a number of the plasmids showing that an entire epimerase gene/protein is not required in order to maintain epimerase activity. The epimerase activity is Ca^{2+} dependent.

Example 1

Purification of mannuronan C-5-epimerase (1), partial amino acid sequencing and synthesis of a mixed DNA probe. The enzyme was isolated from liquid cultures of A. vinelandii essentially as described in [Skjåk-Bræk, G. and Larsen, B. Carbohydrate Research, 103, (1982) 137-140]. The cells were removed by centrifugation and the enzyme was isolated by precipitation with 30% ammoniumsulphate and followed by centrifugation for 20 min. at 10000 rpm. The supernatant was then precipitated with 50% ammonium sulphate (final concentration), and the precipitate after centrifugation was dissolved in 0.05 M imidazole/HCl (pH 6.8) containing 0.34 mM $CaCl_2$ and 0.5 mM dithiothreitol. This crude extract was then desalted on a prepacked column (PD-10) of Sephadex G-25 (Pharmacia) equilibrated with the same buffer. The extract was then applied on an alginate-Sepharose column. Proteins bound by non-specific interactions were eluted with 0.1 M NaCl. The epimerase was eluted as a sharp peak with 0.5 M NaCl.To make the enzyme pure enough for protein sequencing, it was dialyzed against TE-buffer overnight, freezedried, and further purified by SDS-PAGE gel electrophoresis (7.5% polyacrylamide in 25 mM Tris, 192 mM glycine, 0.1% sodium dodecyl sulphate, pH 8.3.) followed by electroblotting (in electrophoresis buffer without sodium dodecyl sulphate) onto a polyvinylidene difluoride membrane, poresize 0.45 μm (Millipore). The membrane was stained with Coomassie brilliant blue and air dried, and the protein with Mw 122 kd was cut out for N-terminal

sequencing on a model 477A protein sequencing apparatus from Applied Biosystems. A DNA oligonucleotide was synthesized on the basis of the amino acid sequence information, and this oligonucleotide was used as a probe for screening of the gene library after end-labelling with ^{32}P by polynucleotide kinase.

Example 2

Isolation of DNA from A. vinelandii and construction of a gene library. A. vinelandii cells were harvested and washed once in 0.9% NaCl. They were then lysed according to [Hansen, J.B. and Olsen, R.H., J. Bacteriol., 135, 227-238, (1978)], and the lysate was extracted twice with phenol and twice with chloroform. Nucleic acids were precipitated with ethanol, and the DNA was collected on a glass rod and dissolved in TE-buffer (10mM Tris, 1mM Na_2EDTA, pH 7.9). Further purification was obtained by CsCl/ethidium bromide density gradient centrifugation. After removal of the ethidium bromide by isopropanol extraction, the DNA solution was dialyzed against TE buffer .

The DNA (molecular size greater than 60 kb) was subjected to partial Sau3AI digestion under conditions maximizing the generation of 15-20 kb fragments. After ethanol precipitation the DNA was dissolved in 40 µl TE buffer to give a concentration of 0.5 µg/µl. The DNA was then dephosphorylated with calf intestine phosphatase, followed by inactivation of the enzyme by incubation at 75°C for 10 minutes in the presence of 10 mN nitrilotriacetic acid. The dephosphorylated DNA was precipitated with ethanol and dissolved in 40 µl 0.1 x TE buffer.

EMBL3 vector DNA was digested with BamHI + EcoRI, followed by an isopropanol precipitation step under conditions leaving the short BamHI/EcoRI oligonucleotides in solution [Frischauf A., Lehrach H., Poustka A. and Murray N., J.

Mol. Biol., 170, 827-842, (1983)]. The Sau3AI-digested and dephosphorylated A. vinelandii DNA (1.75 µg) was then ligated with the BamHI/EcoRI-digested vector DNA (4.75 µg), using T4 DNA ligase in a total reaction volume of 20 µl. After ligations over night at 10°C, 10 µl of the ligation mixture was subjected to in vitro packaging in a Promega Biotech packaging system. The in vitro constructed phage particles were titrated on the E. coli strain Q359, and the library was finally amplified on Q359 in one cycle by plating on solid medium. Screening of the library was performed according to standard protocols [Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular Cloning, Laboratory Manual, 2nd Ed., Cold Spring Harbour Laboratory Press, (1989)], except that the highest stringency wash was 3.2 M tetramethylammoniumchloride at 50°C. A total of 1.4 x 10^5 primary recombinant phages were constructed, a library complexity far above what is required for obtaining representativity of A.vinelandii genes.

Example 3

Measurements of epimerase activity from mannuronan C-5-epimerase (1). (5-^{3}H) alginate was prepared as described in [Skjåk-Bræk, G. and Larsen, B., Carbohydrate Res., 103, 133-136, (1982)]. The (5-^{3}H)alginate was produced by growing Azotobacter vinelandii in a medium consisting of D-Glucose (20g), K_2HPO_4 (1g), $MgSO_4.7\ H_2O$ (200mg), $FeSO_4.\ 7\ H_2O$ (50mg), $NaMoO_4.\ 2\ H_2O$ (5mg), NH_4OAc (2.3g) and $CaCl_2.\ 2\ H_2O$ (50mg) diluted to one litre with water. The cells were grown at 30°C with vigorous shaking. After 30 hours, D-[5-^{3}H]glucose was added to a concentration of 0.6mg/ml (Specific activity, 0.7 µCi/mg) and the cells were allowed to grow for another 72 hours. The culture was cooled in an ice-bath, and the cells were removed by centrifugation. The supernatant solution was dialysed against 0.05M sodium EDTA (3x5 litres) for 24 hours followed by exhaustive dialysis against distilled water. The sodium alginate was then

precipitated with ethanol in the presence of 0.2% of sodium chloride. The specific activity of the label was 29 000 dpm/mg alginate. The composition of this alginate was also analyzed by NMR spectroscopy, and it was found to contain 59% mannuronic acid. Phage lysates were prepared by plating 105 phages per plate. Two ml 2,3,6-trimethylpyridine buffer (50 mM, pH 6.9) were added to each plate, and the softagarose/buffer mixture was scraped off, vortexed and centrifuged at 10 000 rpm for 10 min. The supernatant was used for incubations with (5-^{3}H)alginate by mixing 0.25 ml (5-^{3}H)alginate (2.5 mg/ml), 6,3 µl 0.1 M $CaCl_2$, and 1.45 ml phage lysate. The mixture was incubated at 30°C overnight, and the alginate was precipitated by addition of 15 µl 5 M NaCl and 2 ml ethanol. After incubation at -20°C for 30 min. the solution was centrifuged at 10.000 rpm for 30 min., and 1 ml of the supernatant was used for determination of released ^{3}H [Skjåk-Bræk, G. and Larsen, B., Carbohydrate Res., 103, 133-136, (1982)]. in a liquid scintillation counter.

For measurements of epimerase activity as release of ^{3}H in cells containing recombinant plasmids, the cell cultures were harvested by centrifugation and resuspended in 2,3,6 trimethylpyridine buffer. When IPTG (3 mM) was used for induction of the lac-promoter, the inducer was added to exponentially growing cells and incubations were continued for 3 hours. Cells were disrupted by ultra-sonication, and varying amounts of the lysates were incubated with shaking together with 100 µl (5-^{3}H) alginate (2.5 mg/ml) and 400 µl 2,3,6-trimethylpyridine buffer (total volume 0.6 ml) in the presence of 3.3 mM $CaCl_2$. The quantities of enzyme-containing cell extracts used were adjusted such that the measurements were performed under conditions where the enzyme represented the limiting factor. After incubation at 30°C at the times indicated in each case, the mixtures were precipitated with ethanol under the conditions described above for phage lysates and 1.0 ml of the supernatant was

used for scintillation counting. Controls (using the appropriate host with the pUC18 vector) gave low backgrounds and these numbers were subtracted in the values presented in Table 3.

Example 4

Molecular cloning of a DNA fragment expressing a mannuronan C-5-epimerase activity in E. coli. The A.vinelandii gene library was constructed by cloning partially Sau3AI-digested A.vinelandii DNA into the bacteriophage λ vector EMBL3. In order to identify the epimerase gene in this library, we constructed a DNA probe based on the assumption that the previously purified 122 kd protein represented the epimerase [Skjåk-Bræk, G. and Larsen, B., Carbohydr. Res., 103, 137-149 (1982)]. Initially we tried to use the corresponding protein solution for determination of the N-terminal amino acid sequence of the 122 kd protein, but the results showed that the preparation was not sufficiently pure for this purpose. We therefore purified the protein further by SDS-polyacrylamide gel-electrophoresis, followed by electroblotting onto a membrane. The band containing the 122 kd protein was cut out from this membrane and subjected to N-terminal amino acid sequence analysis. Based on parts of this sequence, we synthesized the mixed DNA probe shown in Figure 1.

The DNA probe synthesized as in Example 1 was labelled with ^{32}P and then used for screening of the A. vinelandii gene library. Clones which hybridized reproducibly against the labelled probe were identified at a frequency of approximately 10^{-3}, and six such clones were selected for further studies. Phage lysates were prepared from each of the six clones, and each lysate was assayed for epimerase activity (Table 2). As can be seen, the lysates prepared from all six clones appeared to contain a weak enzyme activity that could represent the epimerase. This conclusion was further supported by the observation that

control lysates prepared from randomly picked recombinant phages in the library, gave reproducibly lower activity, representing the background activity.

Example 5

Subcloning of a DNA fragments encoding the epimerase. DNA from phage EP2 was partially digested with Sau3AI, and fragments ranging from 4 to 9 kb in size were subcloned in the BamHI site of plasmid pUC18. Cell extracts from DH5α transformants containing recombinant plasmids were assayed for epimerase activity, and the corresponding plasmids were also hybridized against the synthetic oligonucleotide used for screening of the gene library. The analysis of the cell extracts showed that one of them contained an enzymatic activity consistent with the assumption that a polypeptide having epimerase activity was expressed from the plasmid (pHE1) in this clone (see Table 3). We have also tried to centrifuge the extract at 30000 g for 3.5 hours, and found that the activity was not significantly reduced in the supernatant. Since we were unable to detect any significant activity in the culture medium, we conclude that the epimerase is localized intracellularly in E.coli. The insert in pHE1 also hybridized against the synthetic oligonucleotide used for screening, and pHE1 was therefore selected for further analysis.

Example 6

Characterisation of the cloned DNA required for expression of the epimerase, and stability of the enzyme in vivo and in vitro. The insert in pHE1 is approximately 4 kb in size, and Figure 2 shows the restriction map of this insert. Hybridization analysis of pHE1 with the original synthetic oligonucleotide showed that the sequence hybridizing to the oligonucleotide was localized downstream of the SphI site. The hybridizing sequence was further characterized by DNA

sequencing, and this analysis showed that one of the potential reading frames of the sequence was in 100% agreement with the original N-terminal amino acid sequence of the 122 kd protein. Surprisingly, however, the orientation of the sequence was such that it would be transcribed out of the cloned fragment (see Figure 2). This result thus indicated that the observed epimerase activity was not correlated with the sequence encoding the 122 kd protein, a conclusion that was further confirmed by the observation that the terminal 0.5 kb SphI fragment could be deleted (generating plasmid pHE7) without loss of the epimerase activity from the corresponding cell extract. In addition to the SphI deletion, we deleted (from pHE7) the 0.7 kb KpnI fragment at the opposite terminus of the insert, generating plasmid pHE5. As shown in Table 3, pHE5 (in DH5α) expressed the epimerase at a level approximately 27 times higher than the level of expression from pHE1.

During the expression studies described above we found that the measurements were quantitatively difficult to reproduce unless the time of harvesting the cells were kept as constant as possible. We have analyzed this problem more carefully by measuring the enzyme activity at different stages of growth of the E. coli cells. The results of such an analysis are shown in Figure 3, and indicate that the enzymatic activities in the cell extracts are drastically reduced shortly after the cells have entered the stationary phase. To obtain optimal enzyme yields it is therefore important to harvest the cells at the end of the exponential phase or at the beginning of the stationary phase. The reason for the reduction of epimerase activity might potentially be due to proteolysis of the epimerase in stationary phase cells. To study the stability of the enzyme in vitro we have also analyzed the kinetics of ^{3}H release in the DH5α(pHE5) extracts. As can be seen from Figure 4, the enzyme activity is linear over at least 30 hours, demonstrating that the enzyme is very stable in

vitro. A critical parameter for obtaining reproducible results is thus the time of harvesting of the cells.

Example 7

Stimulation of the epimerase activity by induction of the lac promoter. The results described above showed that the levels of expression of the epimerase from pHE5 was significantly higher than in pHE1. The reasons for this could potentially be that the lac promoter was important for the expression, and we have therefore analysed this problem more closely. The analyses were performed in the E. coli strain JM105, a strain which expresses high levels of lac repressor, thus allowing a more repressed state of the promoter under uninduced conditions. When cell extracts prepared from uninduced and induced (with IPTG) cells of JM105(pHE1), a significant stimulation of enzyme activity was observed in the induced cells (Table 3). A similar experiment using JM105(pHE5) showed even greater stimulation of the expression of the epimerase upon addition of IPTG in this case. These experiments thus showed that the lac promoter probably is a key element, although not necessarily the only element, involved in the expression of the epimerase from pHE5. The experiments in addition showed that the direction of transcription is from the KpnI towards the SphI site in the insert. The epimerase gene is therefore transcribed in the same direction as the gene encoding the 122 kd protein, whose N-terminal amino acid sequence was used for the isolation of the cloned DNA.

Preliminary experiments on deleting more DNA from the SphI side of the insert indicated that very little could be deleted without loss of the epimerase activity. At the KpnI side, on the other hand, we found that significant deletions were tolerated. Table 3 shows the results of analysis of expression of the epimerase from a plasmid (pHE8) constructed by deleting the 0.8 kb KpnI/SacII

fragment from pHE5. As can be seen, this deletion resulted in a very strong stimulation of the epimerase activity both in uninduced and induced cells. The expression from pHE8 is presumably based on initiation of translation from the Shine-Dalgarno sequence in the vector (localized between the lac promoter and the polylinker). Similarly, high levels of expression were obtained from pHE22 also due to the coding sequences being in frame with the Shine-Dalgano sequence. So far we have not obtained expression of the epimerase in constructs where deletions have extended beyond the SacII site.

Example 8

Use of a different promotor than the lac promotor. The insert in pHE5 (EcoRI-HindIII) was sub-cloned into plasmid pT7-3 (a derivative of pT7-1 described by [Tabor, S., and C.C. Richardson (1985). Proc. Natl. Acad. Sci. 82, 1074-1078]), and the new plasmid was designated pLB1. The insert in pLB1 is localized downstream of the φ10 promoter in the vector. This promoter is only recognized by the bacteriophage T7 RNA polymerase, and expression of genes downstream of this promoter thereby becomes dependent on expression of this polymerase activity in the cells. The 442 bp (see Figure 2) SacI- SpoI fragment was finally deleted from the insert in pLB1, generating plasmid pLB2. pLB2 was transformed into E. coli K38 (pGP1-2). Plasmid pGP1-2 encodes the gene for T7 RNA polymerase, and the expression of the gene is controlled by a temperature inducible repressor. K38(pLB1, pGP1-2) was grown in exponential phase at 30°C for 4 1/2 hours. One of two parallel cell cultures was then transferred to 42°C for 30 minutes to induce the T7-polymerase. The other parallel cell culture was grown at 30°C for 5 hours. The epimerase activities in the cells were measured as described in example 3, and the results of the measurements are shown in Table 3.

Example 9

Cloning of mannuronan C-5-epimerase (2). Plasmid pHE12 was constructed by inserting a 6.2 kb XhoI fragment from the recombinant bacteriophage lambda derivative EP2 into pUC128. As can be seen from Figure 2 the insert in pHE12 is partly overlapping with the insert in pHE1. Analysis of extracts prepared from cells containing pHE12 (as described for pHE1), showed that they expressed mannuronan C-5-epimerase activity (Table 3). Further analysis showed that the 2.5 kb SpoI-XhoI fragment could be deleted from the insert in pHE12 without affecting the expression of mannuronan C-5-epimerase. Further plasmids were constructed (see Figure 2) and the activity analysed (see Table 3). This demonstrated that both the genes and gene fragments were able to express epimerase activity. The nucleotide sequences of the inserts were determined by the method of Sanger [Sanger,F., S. Nicklen, and A.R. Coulson. 1977. Proc. Natl. Acad. Sci. 74, 5436]. The nucleotide sequences are shown in Figure 6.

Example 10

Sequence Comparison. Five genes have been identified as shown in Figure 2. The insert containing E5 is located about 5-10 kilobases away from the other genes. Figure 6 shows the nucleotide sequence for the complete genes of E4, E1, E2 and a large portion of E3. Detailed analysis of the nucleotide and amino acid sequences revealed highly homologous regions within each gene and between the various genes. Figure 5 characterises each of the genes by reference to the homologous blocks. Each of the genes has at least one A-element and at least one R-element.

E1, E2 and E4 all end with a reasonably homologous sequence termed the S-element (not shown in Figure 5). The last 14 amino acids of the S-element of E1 and E2 are identical with one exception.

Figures 7-10 show detailed analysis of the A- and R-elements within each gene by reference to the consensus sequence (con). Each A-element is approximately 1,150 base pairs long and each R-element is approximately 450 base pairs long. Short oligonucleotides are present in E1, E2 and E3 between the second and third R-elements. Gaps have been introduced where necessary to maximise alignment (see in particular the third R-element of E2).

Hybridization with a probe made from the first part of the A-element to a Southern blot of A. vinelandii digested with restriction endonuclease BglII gave 5 distinct bands. One of these bands contained two A blocks, and another of these bands contained two different fragments with the same size. The number of bands were the same when other strains (ATCC 478, ATCC 12837 and ATCC 12518) of the same species were used. This implies that the bacterium contains at least 5 copies of the A-element, and that this is common to several independently isolated strains of A. vinelandii.

The first part of each R-element contains six perfect and imperfect repeats of a nonapeptide with the consensus sequence LXGGAGXDX, except for the third R-element of E2 which lacks two of these repeats. Figure 12 shows the complete nucleotide and corresponding amino acid sequence of E2. The nonapeptides have been marked with double lines for a good match with the consensus sequence and single lines for less good matches. This nonapeptide motif is characteristic of the haemolysin family of secreted proteins (Suh, Y. and Benedik, M.J., J. Bacteriol 174, (1992) 2361-2366). These proteins are all calcium dependent and are secreted by a pathway which does not involve cleavage of an N terminal signal peptide. For haemolysin secreted from E.coli it has been proposed that the nonamers are responsible for the binding of calcium ions (Ludwig, A. et al, Mol. Gen. Genet. 214, (1988) 553-561, Boehm, D.F. et al, Infect. Immun. 58 (1990) 1959-

1964). It appears that the R-elements are involved in calcium ion binding, calcium being necessary for both enzyme activity and gel formation.

Example 11

Making an altered epimerase.

As can be seen from Table 3, various elements may be deleted from the gene, while maintaining the expression of a protein having epimerase activity. Clearly, the latter portion of E1 having the sequence ARS has epimerase activity (see plasmid pHE8), although deletions in A2 of E1 are not tolerated (see Example 7). Additionally, E2 having the sequence ARRRRS also demonstrates epimerase activity. Additionally, fragments of E3 lacking a carboxy terminal and having the sequences ARRR and ARRRARR have expressed epimerase activity. (See plasmids pH18 and pBD6). Accordingly, it appears that an S-element is not essential for epimerase activity, although the presence of this element may affect activity. We therefore postulated that an epimerase may need at least one A-element and at least one R-element, and that it should be possible to make altered epimerases by combining these elements in different ways. To show this, we constructed a plasmid encoding an epimerase with the sequence RARS:

The insert in pHE1 (EcoRI-HindIII) was subcloned into plasmid pTrc99A (Pharmacia), generating plasmid pHE21. This plasmid contains a trc-promoter in front of the epimerase I gene, a strong transcription termination signal downstream of the gene, and the $lacI^q$-gene allowing induction with IPTG. pHE21 was digested with KpnI and SpoI, made blunt-ended with S1 nuclase and religated. The resulting plasmid, pHE22, expresses a protein having the carboxy terminal of epimerase 1, RARS. The epimerase activity was measured as in Example 3, see Table 3.

Given that epimerase activity is expressed from a number of the constructs, it seems likely that a number of synthetic enzymes may be produced having epimerase activity including differing numbers of A, R and S blocks. The presence of activity in pHE22 implies that it is not essential to have an amino-terminal A block, and so block order may also be altered.

Example 12

The ^{1}H-NMR spectra of alginate epimerased by extracts from plasmids pHE8 and pBD9 show that the proteins encoded by these plasmids have different enzyme activity, pHE8 producing epimerase with single G activity while pBD9 producing epimerase with G block activity. pHE8 encodes the carboxy terminal ARS of E1 whereas pBD9 encodes ARRRRS of E2. The naturally encoded epimerases may therefore have differing activity particularly in the distribution patterns of Gs. The different activity of the various epimerases encoded within the 5 genes could be used to create alignates having a desired structure, by selectively expressing a desired gene or genes. Alternatively, it may be possible to construct synthetic enzymes varying the A, R and S block content of each epimerase to provide enzymes having altered activity providing a further level of control in the production of desired alginates.

Table 1-1 Bacterial strains, phages, and plasmids.

Strain/phage/ plasmid	Remarks	References
Bacterial strains		
A. vinelandii	Strain E	Larsen and Haug (1971)
E. coli		
Q359	supE hsdR $\phi 80^r$P2	Karn et al. (1980)
DH5α	supE44ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1	Bethesda Research Laboratories (1986)
JM105	supE endA sbcB15 hsdR4 rpsL thiΔ(lac-proAB)	Yanisch-Perron et al (1985)
JM109	recA1 supE44 endA1 hsdR17 gyrA96 relA1 thiΔ(lac-proAB) F[traD36 proAB $^+$lacI^{q}lacZΔM15]	Yanisch-Perron et al (1985)
SURE™	e14$^-$(mcrA), Δ(mcrCB-hsdSMR-mrr) 171, endA1, supE44, thi-1, gyrA96, relA1, lac, recB, recJ, sbcC, umuC:Tn5(kan^r), uvrC, [F' proAB, lacI^{q}ZΔM15, Tn10, (tet^r)].	Greener (1990)
Phages		
EMBL3	Bacteriophage λ vector used for construction of A. vinelandii gene library	Frischauf et al. (1983)
EPx	Randomly picked phage from A. vinelandii gene library	See examples
EP2, -3, -6, -7, -8 and -9	Phages isolated from A. vinelandii gene library Identified by hybridization and expresses mannuronan C-5-epimerase	See examples
Plasmids		
pUC18	Ampicillin resistance, ColE1 replicon	Norrander et al.
pUC128	Ampicillin resistance, ColE1 replicon	Keen et al. (1988)
pTrc99A	Ampicillin resistance, ColE1 replicon	Pharmacia
pT7-3	Ampicillin resistance, ColE1 replicon	Tabor & Richardson (1985)
pGP1-2	Kanamycin resistance, P15A replicon	Tabor & Richardson (1985)
pBluescript II SK(+)	Ampicillin resistance, ColE1 replicon	
pHE1	Derivative of pUC18 where a 4 kb Sau3A1 DNA fragment from phage EP2 was subcloned into the BamH1 polylinker	See examples

Table 1-2

pHE7	Derivative of pHE1 where a 0.5 kb Sph1 DNA fragment was deleted	See examples
pHE5	Derivative of pHE7 where a 0.7 kb Kpn1 DNA fragment was deleted	See examples
pHE8	Derivative of pHE5 where a 0.8 kb Kpn1/Sac I DNA fragment was deleted. Cohesive ends were removed prior to ligation by using the 3' exonuclease activity of T4 DNA polymerase	See examples
pLB1	Derivative of pT7-3, where the 2,7 kb insert from pHE5 was cloned into the EcoRI/HindIII polylinker	See examples
pLB2	Derivative of pLB1, where a 0.4 kb SacI/SpoI fragment was deleted	See examples
pHE12	Derivative of pUC128, where a 6.2 kb XhoI fragment from phage EP2 was subcloned into the XhoI polylinker	See examples
pBD1	Derivative of pHE12, where a 2.0 kb SpoI/NsiI fragment was deleted	See examples
pHE21	Derivative of pTrc99A, where the 4.0 kb insert of pHE1 was cloned into the EcoRI/HindIII polylinker	See examples
pHE22	Derivative of pHE21, where a 1.2 kb KpnI/SpoI fragment was deleted	See examples
pHE2	Derivative of pUC18 where a 6.0 kb SphI DNA fragment from phage EP6 was cloned into the SphI site in the polylinker of the vector	See examples
pHE16	Derivative of pHE2 where a 1.5 kb EcoRI-SmaI DNA fragment was deleted	See examples
pBD9	Derivative of pBD1 where a 0.4 kb XhoI-FspI DNA fragment was deleted	See examples
pBD6	Derivative of pHE12 where a 3.4 kb XhoI-EspI DNA fragment was deleted	See examples
pHE18	Derivative of pUC128 where an 5.1 kb NotI-PvuII DNA fragment from EP6 was cloned into the NotI-EcoRV sites in the polylinker of the vector	See examples
pHE14	Derivative of pUC128 where a 3.0 kb BglII DNA fragment from EP6 was cloned into the BamHI site of the polylinker	See examples
pML1	Derivative of pBluescript II SK(+) where a 4.3 kb KpnI-SacII DNA fragment was cloned into the corresponding sites in the polylinker	See examples

References to Table 1

Larsen, B., and A. Haug. 1971. Biosynthesis of alginate. Carbohydr. Res. 17:287-296.

Karn, J., S. Brenner., L. Barnett, and G. Cesareni. 1980. Proc. Natl. Acad. Sci. U.S.A. 77:5172.

Bethesda Research Laboratories. 1986. Bethesda Res. Lab. Focus 8(2):9.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Gene 33:103-119.

Frischauf, A., H. Lehrach, A. Poustka, and N. Murray. 1983. J. Mol. Biol. 170:827-842.

Norrander, J., T. Kempe, J. Messing. 1983. Gene 26:101-106.

Keen N.T, S. Tamaki, D. Kobayashi and D. Trollinger (1988). Gene 79: 191-197.

Tabor S. and C.C. Richardson (1985). Proc Natl Acad Sci USA 82: 10741078.

Greener, A. (1990) Strategies 3: 5-6.

Table 2. Putative mannuronan C-5-epimerase activity in recombinant phage lysates.

Recombinant phage	^3H release (dpm)
EPx	39
EP2	91
EP3	107
EP6	74
EP7	75
EP8	245
EP9	75

EPx originated from the A.vinelandii gene library as a randomly picked plaque, while the other six phages were selected on the basis of the hybridization between their DNA and the labelled oligonucleotide used for screening of the library.

Table 3. Activity of the mannuronan C-5-epimerase expressed from the plasmids.

Enzyme	Strain	Released $^3H/OD_{600}$ unit cell culture	
		No IPTG	IPTG
Epimerase 1^1	JM109(pHE16)	10000	110000
Epimerase 1^2	DH5α(pHE1)	273	nd
Epimerase 1^3	DH5α(pHE5)	9700	nd
Epimerase 1^4	JM105(pHE1)	637	2800
Epimerase 1^5	JM105(pHE5)	4800	28500
Epimerase 1^6	JM105(pHE8)	58900	181000
Epimerase 1^7	JM109(pHE21)	93	1283
Epimerase 1^8	JM109(pHE22)	5383	34611
Epimerase 1^3	K38(pGP1-2,pLB2)*	2150	8333
Epimerase 2^1	JM109(pHE12)	nd	140
Epimerase 2^2	JM109(pBD9)	nd	6700
Epimerase 3^1	JM109(pHE18)	551	2270
Epimerase 3^2	JM109(pBD6)	nd	530
Epimerase 4	DH5α(pHE14)	nd	3500

The extracts were incubated with the alginate for 16 hours, and the numbers are given in dpm. nd = not determined.
* The culture was not induced by IPTG, but by raising the temperature from 30°C to 42°C.

Table 4. NMR analysis of the reaction product after incubation of the recombinant epimerase with a low guluronic acid containing substrate.

Strain	$CaCl_2$ (mM)	Frequences of M and G residues F_M	F_G	F_{MM}	F_{MG}	F_{GG}	F_{GG}/F_G	
JM105(pHE5)	0.85	0.80	0.20	0.66	0.14	0.06	0.3	
JM105(pHE5)	3.4	0.71	0.29	0.45	0.26	0.03	0.10	
DH5α(pHE8)	3.4	0.72	0.28	0.52	0.20	0.08	0.29	
JM109(pHE16)*	3.4	0.82	0.18	0.79	0.03	0.15	0.83	
JM109(pBD9)	2.1	0.60	0.40	0.54	0.06	0.34	0.85	

*This spectrum was obtained at a 400 MHz instrument to be able to get relatively correct figures in spite of the lower conversion.

SEQUENCE LISTING

SEQUENCE LISTING NO: 1
SEQUID NO:1
SEQUENCE TYPE: Nucleotide with corresponding proteins.
SEQUENCE LENGTH: 12411 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE ORGANISM: *Azotobacter vinelandii* strain E

FEATURES:
from 290 to 1951 bp epimerase 4
from 2227 to 6438 bp epimerase 1
from 6702 to 9695 bp epimerase 2
from 9973 to 12411 bp first part of epimerase 3

PROPERTIES: *Azotobacter vinelandii* mannuronan C-5-epimerase genes

```
  1  GATCCGGCCG TCTGAGACGG CGCCTCCGGC CGTCGGCGAG TGCGCCGTTC GCCGACGGCC
 61  GGGCGAACGG ATGAGGACTG CTCCACTCTC ACCCAGATAA GCGCGTGGGC CGTTTCATCC
121  GAGCGCCTTT CCGGGCCGCT TCGAAAGACC GCCACGAGGC ACTCTGTGCA AGGGCCAGGC
181  AGTCGCGTTG CAACCGGAGA CGGGACCGGC CCGTTCGGGC GTCGTCTCTT CCCGCTCCAC
241  TTTTTCCAGG CAGCTTCGGC TGCTCCACTC GGAACCGGGA AGCGGAGAT

290  ATG GAT TAC AAC GTC AAG GAT TTC GGT GCA TTG GGC GAC GGC GTC AGC
     Met Asp Tyr Asn Val Lys Asp Phe Gly Ala Leu Gly Asp Gly Val Ser

338  GAC GAC CGG GCC TCC ATC CAG GCG GCG ATC GAT GCC GCC TAC GCC GCC
     Asp Asp Arg Ala Ser Ile Gln Ala Ala Ile Asp Ala Ala Tyr Ala Ala

386  GGT GGC GGT ACC GTC TAC CTG CCG GCC GGC GAG TAC CGG GTC AGC GCC
     Gly Gly Gly Thr Val Tyr Leu Pro Ala Gly Glu Tyr Arg Val Ser Ala

434  GCC GGG GAG CCG GGC GAC GGC TGC CTG ATG CTC AAG GAC GGC GTC TAC
     Ala Gly Glu Pro Gly Asp Gly Cys Leu Met Leu Lys Asp Gly Val Tyr

482  CTG GCC GGT GCC GGC ATG GGC GAG ACG GTG ATC AAG CTG ATC GAC GGC
     Leu Ala Gly Ala Gly Met Gly Glu Thr Val Ile Lys Leu Ile Asp Gly

530  TCC GAC CAG AAG ATC ACC GGC ATG GTC CGC TCG GCC TAC GGC GAG GAA
     Ser Asp Gln Lys Ile Thr Gly Met Val Arg Ser Ala Tyr Gly Glu Glu

578  ACC AGC AAC TTC GGC ATG CGC GAC CTG ACC CTC GAC GGC AAC CGC GAC
     Thr Ser Asn Phe Gly Met Arg Asp Leu Thr Leu Asp Gly Asn Arg Asp

626  AAC ACC AGC GGC AAG GTC GAC GGC TGG TTC AAC GGC TAT ATC CCC GGC
     Asn Thr Ser Gly Lys Val Asp Gly Trp Phe Asn Gly Tyr Ile Pro Gly

674  GGG GAC GGC GCC GAC CGC GAC GTG ACC ATC GAG CGG GTG GAG GTC CGC
     Gly Asp Gly Ala Asp Arg Asp Val Thr Ile Glu Arg Val Glu Val Arg

722  GAG ATG TCC GGC TAC GGC TTC GAC CCC CAC GAG CAG ACC ATC AAC CTG
     Glu Met Ser Gly Tyr Gly Phe Asp Pro His Glu Gln Thr Ile Asn Leu

770  ACG ATC CGC GAC AGC GTG GCC CAC GAC AAC GGC CTC GAC GGC TTC GTC
     Thr Ile Arg Asp Ser Val Ala His Asp Asn Gly Leu Asp Gly Phe Val

818  GCC GAC TAC CTG GTC GAC AGC GTG TTC GAG AAC AAC GTC GCC TAC GCC
     Ala Asp Tyr Leu Val Asp Ser Val Phe Glu Asn Asn Val Ala Tyr Ala

866  AAC GAC CGC CAC GGC TTC AAC GTG GTC ACC AGC ACC CAC GAT TTC GTC
     Asn Asp Arg His Gly Phe Asn Val Val Thr Ser Thr His Asp Phe Val

914  ATG ACC AAC AAC GTC GCC TAC GGC AAC GGC AGC AGC GGC CTG GTG GTG
     Met Thr Asn Asn Val Ala Tyr Gly Asn Gly Ser Ser Gly Leu Val Val
```

```
 962  CAG CGG GGT CTG GAG GAC CTC GCG CTG CCC AGC AAC ATC CTG ATC GAC
      Gln Arg Gly Leu Glu Asp Leu Ala Leu Pro Ser Asn Ile Leu Ile Asp

1010  GGC GGC GCC TAC TAC GAC AAC GCC CGC GAA GGC GTG CTG CTC AAG ATG
      Gly Gly Ala Tyr Tyr Asp Asn Ala Arg Glu Gly Val Leu Leu Lys Met

1058  ACC AGC GAC ATC ACC CTG CAG AAC GCC GAT ATC CAC GGC AAC GGC TCC
      Thr Ser Asp Ile Thr Leu Gln Asn Ala Asp Ile His Gly Asn Gly Ser

1106  TCC GGG GTG CGC GTC TAC GGC GCC CAG GAC GTG CAG ATC CTC GAT AAC
      Ser Gly Val Arg Val Tyr Gly Ala Gln Asp Val Gln Ile Leu Asp Asn

1154  CAG ATC CAC GAC AAC GCG CAG GCG GCC GCC GTG CCC GAG GTC CTG CTG
      Gln Ile His Asp Asn Ala Gln Ala Ala Ala Val Pro Glu Val Leu Leu

1202  CAG TCC TTC GAC GAT ACC GCC GGG GCG TCC GGC ACC TAC TAC ACG ACC
      Gln Ser Phe Asp Asp Thr Ala Gly Ala Ser Gly Thr Tyr Tyr Thr Thr

1250  CTG AAC ACC CGG ATC GAG GGC AAC ACC ATC AGC GGC TCG GCC AAC TCC
      Leu Asn Thr Arg Ile Glu Gly Asn Thr Ile Ser Gly Ser Ala Asn Ser

1298  ACC TAC GGC ATC CAG GAG CGC AAC GAC GGC ACC GAC TAC AGC AGC CTG
      Thr Tyr Gly Ile Gln Glu Arg Asn Asp Gly Thr Asp Tyr Ser Ser Leu

1346  ATC GAC AAC GAC ATC GCC GGG GTG CAA CAG CCC ATC CAA CTG TAC GGA
      Ile Asp Asn Asp Ile Ala Gly Val Gln Gln Pro Ile Gln Leu Tyr Gly

1394  CCT CAC TCG ACG GTA TCC GGC GAA CCC GGC GCG ACA CCG CAA CAG CCG
      Pro His Ser Thr Val Ser Gly Glu Pro Gly Ala Thr Pro Gln Gln Pro

1442  TCC ACG GGA AGC GAC GGC GAG CCA CTG GTC GGC GGC GAC ACG GAC GAC
      Ser Thr Gly Ser Asp Gly Glu Pro Leu Val Gly Gly Asp Thr Asp Asp

1490  CAG CTC CAG GGC GGC TCC GGC GCC GAT CGC CTG GAC GGC GGG GCC GGC
      Gln Leu Gln Gly Gly Ser Gly Ala Asp Arg Leu Asp Gly Gly Ala Gly

1538  GAC GAC ATC CTC GAC GGC GGC GCC GGG CGC GAC CGG CTG AGC GGC GGC
      Asp Asp Ile Leu Asp Gly Gly Ala Gly Arg Asp Arg Leu Ser Gly Gly

1586  GCG GGC GCC GAC ACC TTC GTG TTC TCC GCC CGC GAG GAC AGC TAC CGT
      Ala Gly Ala Asp Thr Phe Val Phe Ser Ala Arg Glu Asp Ser Tyr Arg

1634  ACC GAC ACG GCG GTG TTC AAC GAC CTG ATC CTC GAC TTC GAG GCC AGC
      Thr Asp Thr Ala Val Phe Asn Asp Leu Ile Leu Asp Phe Glu Ala Ser

1682  GAG GAT CGC ATC GAC CTG TCC GCG CTG GGC TTT TCC GGC CTG GGC GAC
      Glu Asp Arg Ile Asp Leu Ser Ala Leu Gly Phe Ser Gly Leu Gly Asp

1730  GGC TAT GGC GGC ACC CTG CTC CTG AAG ACC AAC GCC GAG GGC ACG CGC
      Gly Tyr Gly Gly Thr Leu Leu Leu Lys Thr Asn Ala Glu Gly Thr Arg

1778  ACC TAC CTG AAA AGC TTC GAG GCG GAT GCC GAG GGA CGG CGC TTC GAG
      Thr Tyr Leu Lys Ser Phe Glu Ala Asp Ala Glu Gly Arg Arg Phe Glu

1826  GTC GCC CTG GAC GGC GAC CAC ACG GGC GAT CTT TCC GCC GCC AAT GTG
      Val Ala Leu Asp Gly Asp His Thr Gly Asp Leu Ser Ala Ala Asn Val

1874  GTC TTC GCC GCG ACC GGG ACG ACC ACC GAA CTC GAA GTC CTC GGC GAC
      Val Phe Ala Ala Thr Gly Thr Thr Thr Glu Leu Glu Val Leu Gly Asp

1922  AGC GGC ACG CAG GCC GGG GCG ATC GTC TAG CGCGTCCCGC TCCGACACAT
      Ser Gly Thr Gln Ala Gly Ala Ile Val ---

1972  AGCCGGTCGT CGGCAAGGCG GCCGGCCGCC GGCTGCCCCG AAGTTTCCAA TCTAATCTCA
2032  CCTACAGACA GGCGCGTTCC GGTGCGCCCG AGCGCCGCCC CCGGGAACGA CCGGCAGGGC
2092  GTGTTTGTGC GCAAGGTGCA GGCGGTCGCG CTCGAAGCCA GAGGCAGGGA AAACCTTTTC
2152  CGGCAGTCGT CTCTTCCTTC TCCACTTCCC AGGCAGCCCT GGGCCGAGCA ACACGACGGG

2212  ATTAGGAAGC GGATC ATG GAT TAC AAC GTC AAG GAT TTC GGA GCA
```

Met Asp Tyr Asn Val Lys Asp Phe Gly Ala

2257 CTG GGC GAT GGC GTC AGC GAC GAC ACG GCG GCC ATC CAG GCG GCG ATC
Leu Gly Asp Gly Val Ser Asp Asp Thr Ala Ala Ile Gln Ala Ala Ile

2305 GAC GCC GCC CAC GCG GCG GGC GGC GGC ACC GTC TAC CTG CCG GCC GGC
Asp Ala Ala His Ala Ala Gly Gly Gly Thr Val Tyr Leu Pro Ala Gly

2353 GAA TAT CGG GTC AGC GGC GGC GAG GAG CCT TCC GAT GGT TGT CTG ACC
Glu Tyr Arg Val Ser Gly Gly Glu Glu Pro Ser Asp Gly Cys Leu Thr

2401 ATC AAG AGC AAC GTC CAT ATC GTC GGC GCC GGG ATG GGC GAG ACG GTG
Ile Lys Ser Asn Val His Ile Val Gly Ala Gly Met Gly Glu Thr Val

2449 ATC AAG ATG GTC GAC GGC TGG ACG CAG AAC GTC ACC GGC ATG GTG CGC
Ile Lys Met Val Asp Gly Trp Thr Gln Asn Val Thr Gly Met Val Arg

2497 TCG GCC TAC GGC GAG GAA ACC AGC AAC TTC GGC ATG AGC GAC CTG ACC
Ser Ala Tyr Gly Glu Glu Thr Ser Asn Phe Gly Met Ser Asp Leu Thr

2545 CTC GAC GGC AAC CGC GAC AAC CTG TCC GCC AAG GTC GAC GGC TGG TTC
Leu Asp Gly Asn Arg Asp Asn Leu Ser Ala Lys Val Asp Gly Trp Phe

2593 AAC GGC TAC ATC CCC GGC CAG GAC GGC GCC GAT CGC GAC GTG ACC CTG
Asn Gly Tyr Ile Pro Gly Gln Asp Gly Ala Asp Arg Asp Val Thr Leu

2641 GAG CGG GTG GAA ATC CGC GAG ATG TCC GGC TAC GGT TTC GAC CCC CAC
Glu Arg Val Glu Ile Arg Glu Met Ser Gly Tyr Gly Phe Asp Pro His

2689 GAG CAG ACC ATC AAC CTG ACG ATC CGC GAC AGC GTG GCC CAC GAC AAC
Glu Gln Thr Ile Asn Leu Thr Ile Arg Asp Ser Val Ala His Asp Asn

2737 AGC CTC GAC GGC TTC GTC GCC GAC TAC CAG GTC GGC GGG GTG TTC GAG
Ser Leu Asp Gly Phe Val Ala Asp Tyr Gln Val Gly Gly Val Phe Glu

2785 AAC AAC GTC TCG TAC AAC AAC GAC CGC CAC GGC TTC AAC ATC GTC ACC
Asn Asn Val Ser Tyr Asn Asn Asp Arg His Gly Phe Asn Ile Val Thr

2833 AGC ACC AAC GAC TTC GTC CTG AGC AAC AAC GTC GCC TAC GGC AAC GGC
Ser Thr Asn Asp Phe Val Leu Ser Asn Asn Val Ala Tyr Gly Asn Gly

2881 GGC GCC GGC CTG GTG GTG CAG CGC GGC TCG TAC GAC CTG CCC CAT CCC
Gly Ala Gly Leu Val Val Gln Arg Gly Ser Tyr Asp Leu Pro His Pro

2929 TAC GAC ATC CTG ATC GAC GGC GGC GCC TAC TAC GAC AAC GCC TTG GAA
Tyr Asp Ile Leu Ile Asp Gly Gly Ala Tyr Tyr Asp Asn Ala Leu Glu

2977 GGC GTG CAG CTC AAG ATG GCC CAC GAC GTC ACC CTG CAG AAC GCC GAG
Gly Val Gln Leu Lys Met Ala His Asp Val Thr Leu Gln Asn Ala Glu

3025 ATC TAC GGC AAC GGC CTG TAC GGG GTG CGC GTC TAC GGC GCC CAG GAC
Ile Tyr Gly Asn Gly Leu Tyr Gly Val Arg Val Tyr Gly Ala Gln Asp

3073 GTG CAG ATC CTC GAC AAC CAG ATC CAC GAC AAT TCG CAG AAC GGC GCC
Val Gln Ile Leu Asp Asn Gln Ile His Asp Asn Ser Gln Asn Gly Ala

3121 TAT GCC GAA GTC CTG CTG CAC TCC TAC GAC GAC ACC GCC GGG GTG TCC
Tyr Ala Glu Val Leu Leu Gln Ser Tyr Asp Asp Thr Ala Gly Val Ser

3169 GGC AAC TTT TAC GTC ACC ACC GGC ACC TGG CTC GAA GGC AAC GTC ATC
Gly Asn Phe Tyr Val Thr Thr Gly Thr Trp Leu Glu Gly Asn Val Ile

3217 AGC GGC TCG GCC AAT TCC ACC TAC GGC ATC CAG GAG CGC GCC GAC GGC
Ser Gly Ser Ala Asn Ser Thr Tyr Gly Ile Gln Glu Arg Ala Asp Gly

3265 ACC GAC TAC AGC AGC CTC TAC GCC AAC AGC ATC GAC GGT GTG CAG ACC
Thr Asp Tyr Ser Ser Leu Tyr Ala Asn Ser Ile Asp Gly Val Gln Thr

3313 GGG GCG GTA CGG CTG TAT GGC GCC AAC TCG ACG GTT TCC AGC CAG TCC

```
        Gly Ala Val Arg Leu Tyr Gly Ala Asn Ser Thr Val Ser Ser Gln Ser

3361    GGC AGT GGC CAG CAG GCG ACC CTC GAA GGC AGC GCG GGC AAC GAT GCG
        Gly Ser Gly Gln Gln Ala Thr Leu Glu Gly Ser Ala Gly Asn Asp Ala

3409    CTG AGC GGG ACC GAG GCC CAC GAG ACG CTG CTC GGC CAG GCC GGC GAC
        Leu Ser Gly Thr Glu Ala His Glu Thr Leu Leu Gly Gln Ala Gly Asp

3457    GAC CGC CTG AAC GCC GAT GCC GGC AAC GAC ATC CTC GAC GGC GGG GCA
        Asp Arg Leu Asn Gly Asp Ala Gly Asn Asp Ile Leu Asp Gly Gly Ala

3505    GGG CGC GAC AAC CTG ACC GGC GGC GCG GGC GCC GAC ACC TTC CGC TTC
        Gly Arg Asp Asn Leu Thr Gly Gly Ala Gly Ala Asp Thr Phe Arg Phe

3553    TCC GCG CGC ACC GAC AGC TAC CGC ACC GAC AGC GCC AGC TTC AAC GAC
        Ser Ala Arg Thr Asp Ser Tyr Arg Thr Asp Ser Ala Ser Phe Asn Asp

3601    CTG ATC ACC GAC TTC GAC GCC GAC GAG GAC AGC ATC GAC CTG TCC GCG
        Leu Ile Thr Asp Phe Asp Ala Asp Glu Asp Ser Ile Asp Leu Ser Ala

3649    CTG GGC TTC ACC GGC CTG GGC GAC GGC TAC AAT GGC ACC CTG CTG CTG
        Leu Gly Phe Thr Gly Leu Gly Asp Gly Tyr Asn Gly Thr Leu Leu Leu

3697    AAG ACC AAC GCC GAG GGT ACG CGC ACC TAC CTG AAG AGC TAC GAA GCG
        Lys Thr Asn Ala Glu Gly Thr Arg Thr Tyr Leu Lys Ser Tyr Glu Ala

3745    GAC GCC CAG GGC CGG CGC TTC GAG ATC GCC CTG GAC GGC AAC TTC ACC
        Asp Ala Gln Gly Arg Arg Phe Glu Ile Ala Leu Asp Gly Asn Phe Thr

3793    GGT CTG TTC AAC GAC AAC AAC CTG TTG TTC GAC GCC GCT CCG GCC ACC
        Gly Leu Phe Asn Asp Asn Asn Leu Leu Phe Asp Ala Ala Pro Ala Thr

3841    GGT ACC GAG GGC AGC GAC AAC CTG CTC GGC ACC GAC GCC GGG GAA ACC
        Gly Thr Glu Gly Ser Asp Asn Leu Leu Gly Thr Asp Ala Gly Glu Thr

3889    CTC CTG GGC TAC GGC GGC AAC GAC ACC CTC AAC GGC GGG GCC GGC GAC
        Leu Leu Gly Tyr Gly Gly Asn Asp Thr Leu Asn Gly Gly Ala Gly Asp

3937    GAC ATC CTG GTC GGC GGC GCC GGG CGC GAC AGC CTG ACC GGC GGC GCC
        Asp Ile Leu Val Gly Gly Ala Gly Arg Asp Ser Leu Thr Gly Gly Ala

3985    GGG GCG GAC GTG TTC CGC TTC GAC GCG CTG TCC GAC AGC CAG CGC AAC
        Gly Ala Asp Val Phe Arg Phe Asp Ala Leu Ser Asp Ser Gln Arg Asn

4033    TAC ACC ACC GGC GAC AAC CAG GCC GAC CGC ATT CTC GAC TTC GAC CCG
        Tyr Thr Thr Gly Asp Asn Gln Ala Asp Arg Ile Leu Asp Phe Asp Pro

4081    ACC CTG GAC AGG ATC GAC GTG TCG GCG CTG GGC TTC ACC GGG CTG GGC
        Thr Leu Asp Arg Ile Asp Val Ser Ala Leu Gly Phe Thr Gly Leu Gly

4129    AAC GGC CGC AAC GGC ACC CTC GCC GTG GTG CTC AAC AGC GCC GGC GAC
        Asn Gly Arg Asn Gly Thr Leu Ala Val Val Leu Asn Ser Ala Gly Asp

4177    CGC ACC GAT CTG AAG AGC TAC GAC ACC GAC GCC AAC GGC TAC AGC TTC
        Arg Thr Asp Leu Lys Ser Tyr Asp Thr Asp Ala Asn Gly Tyr Ser Phe

4225    GAG CTT TCC CTC GCG GGC AAC TAC CAG GGG CAG CTC AGC GCC GAG CAG
        Glu Leu Ser Leu Ala Gly Asn Tyr Gln Gly Gln Leu Ser Ala Glu Gln

4273    TTC GTT TTC GCG ACG TCT CAG GGG GGA CAG ATG ACG ATT ATC GAA GGC
        Phe Val Phe Ala Thr Ser Gln Gly Gly Gln Met Thr Ile Ile Glu Gly

4321    ACC GAC GGC AAC GAT ACC TTG CAG GGC ACC GAG GCC AAC GAG CGG CTC
        Thr Asp Gly Asn Asp Thr Leu Gln Gly Thr Glu Ala Asn Glu Arg Leu

4369    CTC GGC CTG GAC GGC CGG GAC AAC CTG AAC GGC GGC GCC GGC GAC GAC
        Leu Gly Leu Asp Gly Arg Asp Asn Leu Asn Gly Gly Ala Gly Asp Asp

4417    ATC CTC GAC GGC GGA GCG GGG CGC GAC ACC CTG ACC GGC GGC ACG GGG
```

```
      Ile Leu Asp Gly Gly Ala Gly Arg Asp Thr Leu Thr Gly Gly Thr Gly

4465  GCC GAC ACC TTC CTG TTC TCC ACG CGT ACC GAC AGC TAC CGC ACC GAC
      Ala Asp Thr Phe Leu Phe Ser Thr Arg Thr Asp Ser Tyr Arg Thr Asp

4513  AGC GCC AGC TTC AAC GAC CTG ATC ACC GAC TTC GAT CCC ACC CAG GAC
      Ser Ala Ser Phe Asn Asp Leu Ile Thr Asp Phe Asp Pro Thr Gln Asp

4561  CGC ATC GAC CTG TCC GGC CTG GGC TTC AGC GGT TTC GGC AAC GGC TAC
      Arg Ile Asp Leu Ser Gly Leu Gly Phe Ser Gly Phe Gly Asn Gly Tyr

4609  GAC GGC ACC CTG CTG CTG CAG GTC AAC GCC GCG GGC ACC CGC ACC TAC
      Asp Gly Thr Leu Leu Leu Gln Val Asn Ala Ala Gly Thr Arg Thr Tyr

4657  CTG AAG AGT TTC GAG GCC GAT GCC AAC GGC CAG CGC TTC GAG ATC GCC
      Leu Lys Ser Phe Glu Ala Asp Ala Asn Gly Gln Arg Phe Glu Ile Ala

4705  CTG GAC GGC GAC TTC AGC GGC CAA TTG GAC AGC GGC AAC GTG ATC TTC
      Leu Asp Gly Asp Phe Ser Gly Gln Leu Asp Ser Gly Asn Val Ile Phe

4753  GAG CCC GCC GTG TTC AAT GCC AAG GAC TTC GGC GCG CTG GGC GAC GGC
      Glu Pro Ala Val Phe Asn Ala Lys Asp Phe Gly Ala Leu Gly Asp Gly

4801  GCC AGC GAC GAC CGG CCG GCC ATC CAG GCG GCG ATC GAC GCC GCC TAC
      Ala Ser Asp Asp Arg Pro Ala Ile Gln Ala Ala Ile Asp Ala Ala Tyr

4849  GCG GCC GGT GGC GGC ACC GTC TAC CTG CCG GCC GGC GAG TAC CGG GTC
      Ala Ala Gly Gly Gly Thr Val Tyr Leu Pro Ala Gly Glu Tyr Arg Val

4897  AGC CCC ACC GGG GAG CCG GGC GAC GGC TGC CTG ATG CTC AAG GAC GGC
      Ser Pro Thr Gly Glu Pro Gly Asp Gly Cys Leu Met Leu Lys Asp Gly

4945  GTC TAC CTG GCC GGC GAC GGC ATA GGC GAA ACG GTC ATC AAG CTG ATC
      Val Tyr Leu Ala Gly Asp Gly Ile Gly Glu Thr Val Ile Lys Leu Ile

4993  GAC GGC TCC GAC CAG AAG ATC ACC GGC ATG GTG CGC TCG GCC TAT GGC
      Asp Gly Ser Asp Gln Lys Ile Thr Gly Met Val Arg Ser Ala Tyr Gly

5041  GAA GAG ACC AGC AAC TTC GGC ATG AGC GAC CTG ACC CTC GAC GGC AAC
      Glu Glu Thr Ser Asn Phe Gly Met Ser Asp Leu Thr Leu Asp Gly Asn

5089  CGC GAC AAC ACC AGC GGC AAG GTC GAC GGC TGG TTC AAC GGC TAC ATC
      Arg Asp Asn Thr Ser Gly Lys Val Asp Gly Trp Phe Asn Gly Tyr Ile

5137  CCC GGC CAG GAC GGC GCC GAC CGC AAC GTG ACC ATC GAG CGG GTG GAA
      Pro Gly Gln Asp Gly Ala Asp Arg Asn Val Thr Ile Glu Arg Val Glu

5185  ATC CGC GAG ATG TCC GGC TAT GGC TTC GAT CCG CAC GAG CAG ACC ATC
      Ile Arg Glu Met Ser Gly Tyr Gly Phe Asp Pro His Glu Gln Thr Ile

5233  AAC CTG ACG ATC CGC GAC AGC GTG GCC CAC GAC AAC GGC CTC GAC GGC
      Asn Leu Thr Ile Arg Asp Ser Val Ala His Asp Asn Gly Leu Asp Gly

5281  TTC GTC GCC GAC TAC CTG GTC GAC AGC GTG TTC GAG AAC AAC GTC GCC
      Phe Val Ala Asp Tyr Leu Val Asp Ser Val Phe Glu Asn Asn Val Ala

5329  TAC AAC AAC GAC CGC CAC GGC TTC AAC ATC GTC ACC AGC ACC TAC GAT
      Tyr Asn Asn Asp Arg His Gly Phe Asn Ile Val Thr Ser Thr Tyr Asp

5377  TTC GTC ATG ACC AAC AAC GTC GCC TAC GGC AAC GGC GGC GCC GGC CTG
      Phe Val Met Thr Asn Asn Val Ala Tyr Gly Asn Gly Gly Ala Gly Leu

5425  ACG ATC CAG CGG GGC TCG GAG GAC CTG GCC CAG CCG ACC GAT ATC CTG
      Thr Ile Gln Arg Gly Ser Glu Asp Leu Ala Gln Pro Thr Asp Ile Leu

5473  ATC GAC GGC GGC GCC TAC TAC GAC AAC GCC CTG GAA GGC GTG CTG TTC
      Ile Asp Gly Gly Ala Tyr Tyr Asp Asn Ala Leu Glu Gly Val Leu Phe

5521  AAG ATG ACC AAC AAC GTC ACC CTG CAG AAC GCC GAG ATC TAC GGC AAC
```

```
      Lys Met Thr Asn Asn Val Thr Leu Gln Asn Ala Glu Ile Tyr Gly Asn

5569  GGC TCC TCC GGC GTG CGC CTG TAC GGC ACG GAG GAC GTG CAG ATC CTC
      Gly Ser Ser Gly Val Arg Leu Tyr Gly Thr Glu Asp Val Gln Ile Leu

5617  GAC AAC CAG ATC CAC GAC AAT TCG CAG AAC GGC ACC TAT CCG GAA GTC
      Asp Asn Gln Ile His Asp Asn Ser Gln Asn Gly Thr Tyr Pro Glu Val

5665  CTG CTG CAG GCC TTC GAC GAC AGC CAG GTC ACC GGT GAG CTG TAC GAG
      Leu Leu Gln Ala Phe Asp Asp Ser Gln Val Thr Gly Glu Leu Tyr Glu

5713  ACC CTG AAC ACC CGG ATC GAA GGC AAT CTC ATC GAC GCT TCG GAC AAC
      Thr Leu Asn Thr Arg Ile Glu Gly Asn Leu Ile Asp Ala Ser Asp Asn

5761  GCC AAC TAT GCG GTG CGC GAG CGC GAC GAC GGC AGC GAC TAC ACC ACG
      Ala Asn Tyr Ala Val Arg Glu Arg Asp Asp Gly Ser Asp Tyr Thr Thr

5809  CTC GTG GAC AAC GAC ATC AGC GGC GGC CAG GTC GCC TCG GTG CAG CTT
      Leu Val Asp Asn Asp Ile Ser Gly Gly Gln Val Ala Ser Val Gln Leu

5857  TCC GGC GCC CAT TCG AGT CTT TCC GGC GGC ACC GTC GAA GTG CCG CAG
      Ser Gly Ala His Ser Ser Leu Ser Gly Gly Thr Val Glu Val Pro Gln

5905  GGG ACC GAC GGC AAC GAC GTG CTG GTC GGC AGC GAT GCC AAC GAC CAG
      Gly Thr Asp Gly Asn Asp Val Leu Val Gly Ser Asp Ala Asn Asp Gln

5953  CTC TAC GGC GGA GCC GGC GAC GAC CGC CTG GAC GGC GGC GCC GGT GAC
      Leu Tyr Gly Gly Ala Gly Asp Asp Arg Leu Asp Gly Gly Ala Gly Asp

6001  GAC CTG CTC GAC GGC GGA GCG GGG CGC GAC GAC CTG ACC GGC GGC ACG
      Asp Leu Leu Asp Gly Gly Ala Gly Arg Asp Asp Leu Thr Gly Gly Thr

6049  GGT GCC GAC ACC TTC GTG TTC GCC GCG CGT ACC GAT AGC TAC CGC ACC
      Gly Ala Asp Thr Phe Val Phe Ala Ala Arg Thr Asp Ser Tyr Arg Thr

6097  GAC GCG GGG GTG TTC AAC GAC CTG ATC CTC GAC TTC GAC GCC AGC GAG
      Asp Ala Gly Val Phe Asn Asp Leu Ile Leu Asp Phe Asp Ala Ser Glu

6145  GAC CGC ATC GAC CTG TCC GCC CTG GGT TTC AGC GGC TTC GGC GAC GGC
      Asp Arg Ile Asp Leu Ser Ala Leu Gly Phe Ser Gly Phe Gly Asp Gly

6193  TAC AAC GGC ACC CTG CTG GTG CAG CTC AGC AGC GCC GGA ACC CGT ACC
      Tyr Asn Gly Thr Leu Leu Val Gln Leu Ser Ser Ala Gly Thr Arg Thr

6241  TAC CTC AAG AGC TAC GAG GAG GAC CTC GAG GGC CGG CGC TTC GAG GTC
      Tyr Leu Lys Ser Tyr Glu Glu Asp Leu Glu Gly Arg Arg Phe Glu Val

6289  GCC CTG GAC GGC GAC CAC ACG GGC GAT CTT TCC GCC GCC AAT GTG GTT
      Ala Leu Asp Gly Asp His Thr Gly Asp Leu Ser Ala Ala Asn Val Val

6337  TTC GCC GAC GAC GGC TCG GCC GCC GTG GCG AGC AGC GAT CCC GCC GCC
      Phe Ala Asp Asp Gly Ser Ala Ala Val Ala Ser Ser Asp Pro Ala Ala

6385  ACA CAG TTG GAG GTG GTC GGC AGC AGC GGC ACC CAG ACC GAT CAA CTC
      Thr Gln Leu Glu Val Val Gly Ser Ser Gly Thr Gln Thr Asp Gln Leu

6433  GCC TGA TCCGACCCCG CCCATACCCG CCCGGCCATT CCGGCCGGGC GAACCAATGG
      Ala ---

6489  TCTTCAGCCC AGTCTCAGGC ACAGCAGCGC GCGAGCCGCT TCGCTTTGTC CGCCCCCCGC
6549  TTTTCTCGCT GAACGCGACG ATCGCCGGGC GCCGGGGGAAG GGTTCGCCGC ATGCCGAGCC
6609  GGGGACGGGA AAAGCCTGTT CGACCAGTCG ACTCTTCCTC CCTTCACTTT CCAGGCAGCC

6669  TGCGGGCTGC GCAGTAACGG AACAGGAAGC AGC ATG GAT TAC AAC GTC AAA
                                          Met Asp Tyr Asn Val Lys

6720  GAT TTC GGG GCG CTG GGC GAT GGC GTC AGC GAC GAT ACG GCC GCC ATC
      Asp Phe Gly Ala Leu Gly Asp Gly Val Ser Asp Asp Thr Ala Ala Ile
```

```
6768  CAG GCG GCG ATC GAT GCC GCC TAC GCG GCC GGC GGC GGC ACC GTC TAC
      Gln Ala Ala Ile Asp Ala Ala Tyr Ala Ala Gly Gly Gly Thr Val Tyr

6816  CTG CCG GCC GGC GAA TAC CGG GTC AGC GGC GGC GAG GAG CCT TCC GAT
      Leu Pro Ala Gly Glu Tyr Arg Val Ser Gly Gly Glu Glu Pro Ser Asp

6864  GGT TGC CTG ACC ATC AAG AGC AAC GTC CAT ATC GTC GGC GCG GGG ATG
      Gly Cys Leu Thr Ile Lys Ser Asn Val His Ile Val Gly Ala Gly Met

6912  GGC GAG ACG GTC ATC AAG CTG GTC GAC GGC TGG GAT CAG GAC GTC ACC
      Gly Glu Thr Val Ile Lys Leu Val Asp Gly Trp Asp Gln Asp Val Thr

6960  GGC ATC GTC CGC TCG GCC TAC GGC GAG GAG ACC AGC AAC TTC GGC ATG
      Gly Ile Val Arg Ser Ala Tyr Gly Glu Glu Thr Ser Asn Phe Gly Met

7008  AGC GAC CTG ACC CTC GAC GGC AAC CGC GAC AAC ACC AGC GGC AAG GTC
      Ser Asp Leu Thr Leu Asp Gly Asn Arg Asp Asn Thr Ser Gly Lys Val

7056  GAC GGC TGG TTC AAC GGC TAC ATT CCC GGC GAG GAC GGC GCC GAC CGC
      Asp Gly Trp Phe Asn Gly Tyr Ile Pro Gly Glu Asp Gly Ala Asp Arg

7104  GAC GTG ACC CTG GAG CGG GTG GAA ATC CGT GAA ATG TCC GGT TAC GGT
      Asp Val Thr Leu Glu Arg Val Glu Ile Arg Glu Met Ser Gly Tyr Gly

7152  TTC GAT CCG CAC GAG CAG ACC ATC AAC CTG ACG ATC CGC GAC AGC GTG
      Phe Asp Pro His Glu Gln Thr Ile Asn Leu Thr Ile Arg Asp Ser Val

7200  GCC CAC GAC AAC GGC CTC GAC GGC TTC GTC GCC GAT TTC CAG ATC GGC
      Ala His Asp Asn Gly Leu Asp Gly Phe Val Ala Asp Phe Gln Ile Gly

7248  GGG GTG TTC GAG AAC AAC GTC TCG TAC AAC AAC GAC CGC CAC GGC TTC
      Gly Val Phe Glu Asn Asn Val Ser Tyr Asn Asn Asp Arg His Gly Phe

7296  AAC ATC GTC ACC AGC ACC AAC GAC TTC GTC CTG AGC AAC AAC GTC GCC
      Asn Ile Val Thr Ser Thr Asn Asp Phe Val Leu Ser Asn Asn Val Ala

7344  TAC GGC AAC GGC GGC GCC GGC CTG GTG GTG CAG CGC GGC TCG TCC GAC
      Tyr Gly Asn Gly Gly Ala Gly Leu Val Val Gln Arg Gly Ser Ser Asp

7392  GTG GCG CAC CCC TAC GAC ATC CTG ATC GAC GGC GGC GCC TAC TAC GAC
      Val Ala His Pro Tyr Asp Ile Leu Ile Asp Gly Gly Ala Tyr Tyr Asp

7440  AAC GGC CTG GAA GGC GTG CAG ATC AAG ATG GCC CAC GAC GTC ACC CTG
      Asn Gly Leu Glu Gly Val Gln Ile Lys Met Ala His Asp Val Thr Leu

7488  CAG AAC GCC GAG ATC TAC GGC AAC GGC CTA TAC GGG GTG CGC GTC TAC
      Gln Asn Ala Glu Ile Tyr Gly Asn Gly Leu Tyr Gly Val Arg Val Tyr

7536  GGC GCC GAG GAT GTG CAG ATC CTC GAC AAC TAC ATC CAC GAC AAT TCG
      Gly Ala Glu Asp Val Gln Ile Leu Asp Asn Tyr Ile His Asp Asn Ser

7584  CAG AAC GGT TCC TAC GCG GAA ATC CTC CTG CAG TCC TAC GAC GAT ACC
      Gln Asn Gly Ser Tyr Ala Glu Ile Leu Leu Gln Ser Tyr Asp Asp Thr

7632  GCC GGG GTG TCC GGC AAT TTC TAC ACC ACC ACC GGC ACC TGG ATC GAA
      Ala Gly Val Ser Gly Asn Phe Tyr Thr Thr Thr Gly Thr Trp Ile Glu

7680  GGC AAC ACC ATC GTC GGC TCG GCC AAC TCC ACC TAT GGC ATC CAG GAG
      Gly Asn Thr Ile Val Gly Ser Ala Asn Ser Thr Tyr Gly Ile Gln Glu

7728  CGC GAC GAC GGC ACC GAC TAC AGC AGC CTC TAC GCC AAC AGC GTC AGC
      Arg Asp Asp Gly Thr Asp Tyr Ser Ser Leu Tyr Ala Asn Ser Val Ser

7776  AAT GTG CAG AAC GGC TCG GTG CGC CTC TAC GGC GCC AAC TCC GTC GTC
      Asn Val Gln Asn Gly Ser Val Arg Leu Tyr Gly Ala Asn Ser Val Val

7824  TCC GAC CTG CCC GGC ACC GGC CAG CAG GCG ACC CTC GAA GGC ACG GCC
      Ser Asp Leu Pro Gly Thr Gly Gln Gln Ala Thr Leu Glu Gly Thr Ala
```

```
7872  GGC AAC GAC ACG CTT GGC GGC AGC GAC GCC CAC GAG ACG CTG CTC GGG
      Gly Asn Asp Thr Leu Gly Gly Ser Asp Ala His Glu Thr Leu Leu Gly

7920  CTG GAC GGC AAC GAC CGC CTG AAC GGC GGC GCC GGC AAC GAC ATC CTC
      Leu Asp Gly Asn Asp Arg Leu Asn Gly Gly Ala Gly Asn Asp Ile Leu

7968  GAC GGC GGC GCC GGG CGC GAC AAC CTG ACC GGC GGC GCG GGC GCC GAC
      Asp Gly Gly Ala Gly Arg Asp Asn Leu Thr Gly Gly Ala Gly Ala Asp

8016  CTG TTC CGC GTC TCC GCG CGC ACC GAC AGC TAC CGC ACC GAC AGC GCC
      Leu Phe Arg Val Ser Ala Arg Thr Asp Ser Tyr Arg Thr Asp Ser Ala

8064  AGC TTC AAC GAC CTG ATC ACC GAC TTC GAC GCC AGC CAG GAC CGC ATC
      Ser Phe Asn Asp Leu Ile Thr Asp Phe Asp Ala Ser Gln Asp Arg Ile

8112  GAC CTG TCC GCG CTG GGC TTC ACC GGG CTG GGC GAC GGC TAT AAC GGC
      Asp Leu Ser Ala Leu Gly Phe Thr Gly Leu Gly Asp Gly Tyr Asn Gly

8160  ACC CTG CTG CTG CAG GTC AGC GCC GAC GGC AGC CGC ACC TAT CTG AAG
      Thr Leu Leu Leu Gln Val Ser Ala Asp Gly Ser Arg Thr Tyr Leu Lys

8208  AGC CTG GAG GCG GAT GCC GAG GGG CGG CGT TTC GAG ATC GCC CTG GAC
      Ser Leu Glu Ala Asp Ala Glu Gly Arg Arg Phe Glu Ile Ala Leu Asp

8256  GGC AAC TTC GCC GGC CTG CTC GGT GCC GGC AAC CTG CTC TTC GAG CGC
      Gly Asn Phe Ala Gly Leu Leu Gly Ala Gly Asn Leu Leu Phe Glu Arg

8304  ACC GCC ATC GAG GGG GAT GCC GGC GAC AAC GCC CTG CTC GGT ACC TCG
      Thr Ala Ile Glu Gly Asp Ala Gly Asp Asn Ala Leu Leu Gly Thr Ser

8352  GCC GCC GAG ACA TTG CTC GGC CAC GCC GGC AAC GAC ACG CTC GAC GGC
      Ala Ala Glu Thr Leu Leu Gly His Ala Gly Asn Asp Thr Leu Asp Gly

8400  GGG GCC GGC GAC GAC ATC CTG GTC GGC GGC GCC GGG CGC GAC AGC CTC
      Gly Ala Gly Asp Asp Ile Leu Val Gly Gly Ala Gly Arg Asp Ser Leu

8448  ACC GGC GGC GCC GGA GCG GAC GTG TTC CGC TTC GAC GCG CTG TCC GAC
      Thr Gly Gly Ala Gly Ala Asp Val Phe Arg Phe Asp Ala Leu Ser Asp

8496  AGC CAG CGC AAC TAC GAC ATC GGC GAC AAC CAG GGC GAC CGC ATC GCC
      Ser Gln Arg Asn Tyr Asp Ile Gly Asp Asn Gln Gly Asp Arg Ile Ala

8544  GAC TTC GCG GTG GGC GAA GAC AAG CTC GAC GTA TCG GCG CTG GGC TTC
      Asp Phe Ala Val Gly Glu Asp Lys Leu Asp Val Ser Ala Leu Gly Phe

8592  ACC GGG CTG GGC GAC GGC TAC AAC GGC ACC CTC GCC CTG GTG CTC AAC
      Thr Gly Leu Gly Asp Gly Tyr Asn Gly Thr Leu Ala Leu Val Leu Asn

8640  AGC GCC GGC GAC CGC ACC TAC GTG AAA AGC TAC GAG AAC GGC GCC GAC
      Ser Ala Gly Asp Arg Thr Tyr Val Lys Ser Tyr Glu Asn Gly Ala Asp

8688  GGC TAC CGC TTC GAG TTT TCC CTC GAC GGC AAC TAT CTG GAG CTA CTC
      Gly Tyr Arg Phe Glu Phe Ser Leu Asp Gly Asn Tyr Leu Glu Leu Leu

8736  GGC AAC GAG GAT TTC ATC TTC GCC ACG CCC AGC GGC CAG CAA CTC CTC
      Gly Asn Glu Asp Phe Ile Phe Ala Thr Pro Ser Gly Gln Gln Leu Leu

8784  GAA GGC AGC GCC GGC AAC GAC AGC CTG CAG GGC ACG GCC GCC GAC GAG
      Glu Gly Ser Ala Gly Asn Asp Ser Leu Gln Gly Thr Ala Ala Asp Glu

8832  GTG ATC CAC GGC GGC GGC GGG CGC GAC ACG CTG GCC GGA GGG GCC GGG
      Val Ile His Gly Gly Gly Gly Arg Asp Thr Leu Ala Gly Gly Ala Gly

8880  GCC GAC GTG TTC CGC TTT AGC GAA CTG ACC GAC AGC TAC CGA GAC AGT
      Ala Asp Val Phe Arg Phe Ser Glu Leu Thr Asp Ser Tyr Arg Asp Ser

8928  GCC AGC TAT GCC GAT CTG ATC ACT GAC TTC GAT GCC AGC GAG GAT CGT
      Ala Ser Tyr Ala Asp Leu Ile Thr Asp Phe Asp Ala Ser Glu Asp Arg
```

```
8976  ATC GAC CTG TCC GGC CTC GGC TTC AGC GGT CTG GGC AAC GGC TAC GGC
      Ile Asp Leu Ser Gly Leu Gly Phe Ser Gly Leu Gly Asn Gly Tyr Gly

9024  GGT ACC CTG GCG CTG CAG GTG AAC AGC GCC GGT ACC CGC ACC TAC CTG
      Gly Thr Leu Ala Leu Gln Val Asn Ser Ala Gly Thr Arg Thr Tyr Leu

9072  AAG AGC TTC GAG ACC AAC GCC GCC GGC GAG CGT TTC GAG ATC GCC CTG
      Lys Ser Phe Glu Thr Asn Ala Ala Gly Glu Arg Phe Glu Ile Ala Leu

9120  GAC GGC GAC CTG TCC GCG CTC GGC GGG GCC AAC CTG ATC CTC GAC GCG
      Asp Gly Asp Leu Ser Ala Leu Gly Gly Ala Asn Leu Ile Leu Asp Ala

9168  CGT ACC GTA CTG GCG GGC GGC GAC GGC AAC GAC ACG CTT TCC GGC AGC
      Arg Thr Val Leu Ala Gly Gly Asp Gly Asn Asp Thr Leu Ser Gly Ser

9216  AGC GCG GCC GAG GAA CTG CTC GGC GGG GTC GGC AAC GAC AGC CTG GAC
      Ser Ala Ala Glu Glu Leu Leu Gly Gly Val Gly Asn Asp Ser Leu Asp

9264  GGC GGC GCC GGC AAC GAC ATC CTC GAC GGC GGG GCG GGG CGC GAC ACC
      Gly Gly Ala Gly Asn Asp Ile Leu Asp Gly Gly Ala Gly Arg Asp Thr

9312  CTG AGT GGC GGC AGC GGC AGC GAC ATC TTC CGC TTC GGC GGC GCG CTC
      Leu Ser Gly Gly Ser Gly Ser Asp Ile Phe Arg Phe Gly Gly Ala Leu

9360  GAC AGC TTC CGC AAC TAC GCC AGC GGG ACG AAC GGC ACC GAC AGC ATC
      Asp Ser Phe Arg Asn Tyr Ala Ser Gly Thr Asn Gly Thr Asp Ser Ile

9408  ACC GAC TTC ACC CCC GGC GAG GAT CTG ATC GAC CTC TCC GTG CTC GGC
      Thr Asp Phe Thr Pro Gly Glu Asp Leu Ile Asp Leu Ser Val Leu Gly

9456  TAC ACC GGG CTG GGC GAC GGC TAC AAC GGT ACC CTG GCG ATA GTG CTG
      Tyr Thr Gly Leu Gly Asp Gly Tyr Asn Gly Thr Leu Ala Ile Val Leu

9504  AAC GAC GCC GGC ACC AAG ACC TAC CTG AAA AAC CGC GAG AGC GAC GCC
      Asn Asp Ala Gly Thr Lys Thr Tyr Leu Lys Asn Arg Glu Ser Asp Ala

9552  GAA GGC AAC CAG TTC GAG ATC GCC CTG GAG GGC AAC CAC GCC GAC CAG
      Glu Gly Asn Gln Phe Glu Ile Ala Leu Glu Gly Asn His Ala Asp Gln

9600  CTC GAT GCG AGC GAC TTC ATC TTC GCC ACG GCG GCC GCG ACC ACC GGA
      Leu Asp Ala Ser Asp Phe Ile Phe Ala Thr Ala Ala Ala Thr Thr Gly

9648  ATC GAG GTG GTC GGC GGC AGC GGC ACC CAG ACC GAT CAG CTC GCC TGA
      Ile Glu Val Val Gly Gly Ser Gly Thr Gln Thr Asp Gln Leu Ala ---

9696  TCCGACCCCG CCCGCACCCG CCCGGCCATT CCGGCCGGGC GAACCAATGG CCTTTTGATC
9756  AGTCTCAGGC ACAGCAACGT GTGCGCCGCT TCGCTTGTTC GCCCTCCCGG CCTTGTTTCT
9816  CGCTGAAAGC GACGATCGCC GGGGGCGTGC CGGGCGCGAG AAAAGGTTCG CCGTGTGCAA
9876  AGCCGGGGAC GGGAAAAGCC TGTTCAAGTA GTCGACTCTT CCTTCTCCTT TTTCCTAGAC

9936  GGCCTCT TGG CTG AGC ATT AAC GGA ACA GGA AGC AGC ATG GAC TTC AAC
                                                  Met Asp Phe Asn

9985  GTC AAA GAT TTC GGG GCA CTG GGC GAT GGC GCC AGC GAC GAC ACG GCG
      Val Lys Asp Phe Gly Ala Leu Gly Asp Gly Ala Ser Asp Asp Thr Ala

10033 GCC ATC CAG GCG GCG ATC GAT GCC GCC CAC GCG GCG GGC GGC GGC ACC
      Ala Ile Gln Ala Ala Ile Asp Ala Ala His Ala Ala Gly Gly Gly Thr

10081 GTC TAC CTG CCG GCT GGC GAG TAT CGG GTC AGC GGC GGC GAG GAG CCT
      Val Tyr Leu Pro Ala Gly Glu Tyr Arg Val Ser Gly Gly Glu Glu Pro

10129 TCC GAC GGC GCG CTG ACC ATC AAG AGC AAC GTC TAT ATC GTC GGC GCC
      Ser Asp Gly Ala Leu Thr Ile Lys Ser Asn Val Tyr Ile Val Gly Ala

10177 GGG ATG GGC GAG ACG GTG ATC AAG ATG GTC GAC GGC TGG ACG CAG AAC
      Gly Met Gly Glu Thr Val Ile Lys Met Val Asp Gly Trp Thr Gln Asn

10225 GTC ACC GGC ATG GTG CGC TCG GCC TAT GGC GAG GAG ACC AGC AAC TTC
```